Report 1489

602166

2 of 3

177 pgs $3.00

DEPARTMENT OF THE NAVY

DAVID TAYLOR MODEL BASIN

HYDROMECHANICS

○

AERODYNAMICS

○

STRUCTURAL MECHANICS

○

APPLIED MATHEMATICS

LECTURES ON APPLIED MATHEMATICS
THE CALCULUS OF VARIATIONS

by

Francis D. Murnaghan

APPLIED MATHEMATICS LABORATORY

RESEARCH AND DEVELOPMENT REPORT

May 1961 Report 1489

LECTURES ON APPLIED MATHEMATICS

THE CALCULUS OF VARIATIONS

by

Francis D. Murnaghan

May 1961 Report 1489

TABLE OF CONTENTS

FOREWORD

This report contains the lecture notes prepared by Professor Francis D. Murnaghan for use in the Second Annual Lecture Series on Applied Mathematics. This series of lectures, on the Calculus of Variations, was presented by Professor Murnaghan in the spring of 1960, under the sponsorship of the Applied Mathematics Laboratory, David Taylor Model Basin.

Just as in the case of the first lecture series in Applied Mathematics,on the subject of the Laplace Transformation, Professor Murnaghan has made a unique contribution to the instruction of applied mathematics, at a time when progress in this field is so vitally important to our national aims. The Applied Mathematics Laborrtory is proud to make these lectures available in report form.

Harry Polachek
HARRY POLACHEK
Technical Director
Applied Mathematics Laboratory

ABSTRACT

These lectures on applied mathematics are devoted to the Calculus of Variations, with especial attention to the applications of this subject to mechanics and wave-propagation. Our aim has been to give more than a superficial account, treating only extremal or stationary curves, and in order to keep the treatment self-contained and free of formidable mathematical difficulties, we have made the necessary differentiability assumptions which are usually satisfied in the applications of the Calculus of Variations to problems of nature. We have emphasized the parametric treatment since this simplifies the passage from the Lagrangian to the Hamiltonian point of view or vice versa. The distinction between extremal and minimal curves is clearly explained, and Legendre's condition for a minimal curve is proved. Extremal fields and the Hilbert Invariant Integral are treated in some detail, the essential distinction between plane problems and problems in space of three or more dimensions, as far as extremal fields are concerned, being carefully explained. The Weierstrass E-function is treated and the connection between the Calculus of Variations and the Rayleigh quotients and the method of Rayleigh-Ritz, so important in vibration problems, is clearly shown. In the treatment of wave-propagation the analogue for earthquake waves of Snell's law of refraction is given. The lectures close with a short account of multiple integral problems and with a discussion of the useful maximum-minimum principle, which we owe to Courant.

Care has been taken to make the treatment self-contained, and details of the proofs of the basic mathematical theorems are given.

Lectures on Applied Mathematics

The Calculus of Variations

Lecture 1

The Lagrangian Function and the Parametric Integrand

The problems of the calculus of variations which we shall treat in these lectures belong to one or the other of two types. The simplest example of the first of these two types may be stated as follows:

Given two points in a plane, or in 3-dimensional Euclidean space, does there exist a curve of shortest length connecting these two points ; and, if so, is this curve unambiguously determinate? We shall refer to the problems of the calculus of variations which belong to the type to which this problem belongs as problems of Type 1. An example of the second of the two types may be stated as follows:

Given a closed curve in 3-dimensional Euclidean space, does there exist in this space, a surface, having this curve as its boundary, whose area is least and, if so, is this surface unambiguously determinate? We shall refer to the problems of the calculus of variations which belong to the type to which this problem belongs as problems of Type 2.

Problems of the calculus of variations of Type 1 are <u>curve</u> problems where the curves with which we are concerned may be plane curves or space curves or, indeed, curves in a space of any number of dimensions

The problem of the motion of a mechanical system may be conveniently stated as a problem of Type 1 of the calculus of variations. If the mechanical system has n degrees of freedom, $n = 1, 2, \ldots$, we write its generalized coordinates as a column, or $n \times 1$, matrix x, so that the j th coordinate $j = 1, \ldots, n$ is denoted by x^j. The velocity $n \times 1$ matrix x_t is the derivative of x with respect to the time t so that the j th element of x_t is x^j_t, $j = 1, \ldots, n$, and the Lagrangian function $L = T-V$ of the system is a function of the two $n \times 1$ matrices x and x_t, and also. if the mechanical system is nonconservative, of the time t. Here, T is the kinetic energy and V is the potential energy of the mechanical system. We assume that L is a continuous function of the matrices x, x_t, and also, if the system is nonconservative, of t; by the statement that L is a continuous function of the two matrices x and x_t we mean that L is a continuous function of the 2n elements x^j, x^k_t, $j = 1, \ldots, n$, $k = 1, \ldots, n$, of these matrices. If, then, $x = x(t)$, $a \leqslant t \leqslant b$, is any smooth curve C in the n-dimensional coordinate space of the mechanical system, so that x and x_t are, along C, continuous functions of t over the interval $a \leqslant t \leqslant b$, the Lagrangian function L is, along C, a continuous function of t over the interval $a \leqslant t \leqslant b$, and we may consider the integral

$$I = \int_a^b L\,dt$$

This integral is a number whose value depends on and is unambiguously determined by the curve C just as the length of a rectifiable curve in a plane, or in 3-dimensional Euclidean space, depends on and is unambiguously determined by the curve. The motion of the mechanical system is such that its paths, i.e., the curves $x = x(t)$, $a \leqslant t \leqslant b$, which describe this motion are such that the integral $I = \int_a^b L\,dt$ has, when evaluated along any one of these paths, a stationary value (without being, necessarily, a minimum or a maximum). Thus the integral I plays, for the mechanical system, the role played by the length of a rectifiable curve in the introductory example we have given of problems of Type 1 of the calculus of variations, but there is one essential difference: I has, when evaluated along a path of the mechanical system, merely a stationary value while the length integral is actually a minimum when the curve of integration gives it a stationary value. In this connection we observe that the time-coordinate space of a mechanical system is what is called a numerical, rather than a metrical, space; the concept of distance between two of its points is not defined, so that it does not make sense to speak of nearby points of this coordinate space nor of the length of a curve in it. We may, however, endow this coordinate space with a metric and we shall do this by assigning to it the ordinary Euclidean metric in accordance with which the distance between any two points $\begin{pmatrix} t \\ x \end{pmatrix}$ and $\begin{pmatrix} t' \\ x' \end{pmatrix}$ is the

magnitude of the matrix $\begin{pmatrix} t'-t \\ x'-x \end{pmatrix}$.

In order to treat most simply the problem of Type 1 of the calculus of variations which is furnished by the motion of a mechanical system it is convenient, particularly when the system is nonconservative so that the Lagrangian function L involves explicitly not only the two $n \times 1$ matrices x and x_t but also the time variable t, to place this variable on an equal footing with the n elements of the coordinate matrix x. To do this we replace our n-dimensional coordinate space by a $(n + 1)$ - dimensional time-coordinate space. We denote $n + 1$ by N and introduce the $N \times 1$ matrix $X = \begin{pmatrix} t \\ x \end{pmatrix}$ whose first element is t and whose remaining n elements are those of the $n \times 1$ matrix x. A smooth curve in the N-dimensional time-coordinate space is furnished by a formula $X = X(\mathcal{T})$, $\alpha \leqslant \mathcal{T} \leqslant \beta$, where $\mathcal{T}$ is any convenient independent variable, or parameter, of which the $N \times 1$ matrix $X(\mathcal{T})$ is a continuously differentiable function. If we choose $\mathcal{T}$ to be t itself, the first of the N equations implied by the formula $X = X(\mathcal{T})$ is simply $t = \mathcal{T}$ but, in general, this equation will be replaced by $t = t(\mathcal{T})$, where $t(\mathcal{T})$ is either smooth, i.e., possesses a continuous derivative, or is at least piecewise smooth over the interval $\alpha \leqslant \mathcal{T} \leqslant \beta$. By the words piecewise smooth, we mean that $t(\mathcal{T})$, while continuous over $\alpha \leqslant \mathcal{T} \leqslant \beta$, may fail to be differentiable at a finite number of interior points of this interval; at each of these points it possesses a right-hand and a left-hand derivative but these derivatives are not equal. At all points

where t_τ is defined it is, by hypothesis, continuous. We make one further restriction on the function $t = t(\tau)$; namely, we assume that $t_\tau > 0$ at all points where t_τ is defined, and this implies that at each of the finite number of points at which t_τ is possibly undefined the right-hand and left-hand derivatives of t with respect to τ are non-negative since the right-hand derivative, for example, of t at $\tau = \tau_1$, say, is the limit, as $\delta \to 0$ through positive values of $t_\tau(\tau_1 + \delta)$. The reason for this restriction is as follows: We regard t, which was the parameter or independent variable used by Lagrange, as a master or control parameter and we do not wish any other parameter τ to sometimes increase and sometimes decrease as t increases. Thus we do not wish t_τ to change sign over the interval $\alpha \leqslant \tau \leqslant \beta$. We could satisfy this wish by requiring that $t_\tau \leqslant 0$ instead of $t_\tau > 0$, but, since a mere change of sign of τ changes the inequality $t_\tau \leqslant 0$ into $t_\tau \geqslant 0$ there is no real loss of generality in requiring that t_τ be > 0 at all the points of the interval $\alpha \leqslant \tau \leqslant \beta$ at which it is defined. We assume, further, that the number of points, if any, of the interval $\alpha \leqslant \tau \leqslant \beta$ at which $t_\tau = 0$ is finite.

When we pass from the master parameter t to any other allowable parameter τ by means of a formula $t = t(\tau), \alpha \leqslant \tau \leqslant \beta$, the integrand of the integral I is changed from the Lagrangian function L to the product of L by t_τ:

$$I = \int_a^b L\,dt = \int_\alpha^\beta L t_\tau\, d\tau; \quad a = t(\alpha),\ b = t(\beta)$$

We denote this new integrand by F and we refer to F as the parametric integrand, the original integrand L being termed the nonparametric , or Lagrangian, integrand of our problem of Type 1 of the calculus of variations. Under an allowable change of parameter $\tau \to \tau'$ furnished by a formula $\tau = \tau(\tau')$, $\alpha' < \tau' < \beta'$, L remains unaffected while $t_\tau \to t_{\tau'} = t_\tau \tau_{\tau'}$, where $\tau_{\tau'}$ is positive save, possibly, at a finite number of points of the interval $\alpha' < \tau' < \beta'$. Since the value of an integral is insensitive to changes of the integrand at a finite number of points, we shall refer to $\tau_{\tau'}$ as positive (rather than nonnegative). Then the transformation $\tau \to \tau'$ of the independent variable, or parameter, induces the transformation $F \to F' = F\tau_{\tau'}$ of the integrand of our integral I. This change of the integrand is necessary to ensure the invariance, or lack of dependence upon the particular parameter adopted, of the integral I itself. F is a function of the two N x 1 matrices X and X_τ and, under the change of parameter $\tau \to \tau'$, the second of these is multiplied by the positive factor $\tau_{\tau'}$ while the first is insensitive to this change of parameter. Thus, k being any positive number, the parametric integrand $F(X, X_\tau)$ is such that

$$F(X, kX_\tau) = kF(X, X_\tau)$$

We express this important quality of the parametric integrand F by the statement that F is a positively homogeneous function, of

degree 1, of the $N \times 1$ matrix X_τ

Example.

Denoting the rectangular Cartesian coordinates of a point in 3-dimensional Euclidean space by (t, x^1, x^2), the formula which furnishes the arc-length I of any smooth curve in this space is

$$I = \int_a^b \left\{1 + \left(x^1_t\right)^2 + \left(x^2_t\right)^2\right\}^{1/2} dt = \int_a^b L\, dt$$

where $L = \left\{1 + \left(x^1_t\right)^2 + \left(x^2_t\right)^2\right\}^{1/2}$. Under the change of parameter $t \longrightarrow \tau$ this appears in the form $I = \int_\alpha^\beta F\, d\tau$,

where $F = L\, t_\tau = \left\{\left(t_\tau\right)^2 + \left(x^1_\tau\right)^2 + \left(x^2_\tau\right)^2\right\}^{1/2}$. Thus F is the magnitude, $\left(X_\tau^* X_\tau\right)^{1/2}$, of the 3×1 matrix

$$X_\tau = \begin{pmatrix} t \\ x^1 \\ x^2 \end{pmatrix}_\tau = \begin{pmatrix} t_\tau \\ x^1_\tau \\ x^2_\tau \end{pmatrix}.$$

Observe that F, while a positively homogeneous function, of degree 1, of X_τ, is not a homogeneous function, of degree 1, of X_τ; when X_τ is multiplied by a negative number k, F is multiplied by $-k = |k|$.

We shall assume from now on that the Lagrangian function L is a continuously differentiable function of the $N \times 1$ matrix X and of the $n \times 1$ matrix x_t. The derivative of L with respect to the $n \times 1$ velocity matrix x_t is a $1 \times n$ matrix which is known as the Lagrangian momentum matrix and which is denoted by p. Similarly the derivative of the parametric integrand F,

which is a continuously differentiable function of the two $N \times 1$ matrices X and X_τ with respect to the $N \times 1$ matrix X_τ, is a $1 \times N$ matrix P, which we term the parametric momentum matrix. In both of these differentiations the $N \times 1$ matrix X is supposed to be held fixed. The first element t_τ of X_τ appears in both the factors,

$L = L(X, x_t) = L\left(X, \frac{x_\tau}{t_\tau}\right)$ and t_τ, of F and so

$$P_1 = F_{t_\tau} = L_{t_\tau} t_\tau + L$$

Now $L_{t_\tau} = -L_{x_t}\frac{x_\tau}{t_\tau^2} = -\frac{px_t}{t_\tau}$, so that

$$P_1 = L - px_t$$

The remaining elements of X_τ, namely, the elements of x_τ, appear only in the first factor L of F and so

$$P_j = L_{x_\tau^{j-1}} t_\tau, \quad j = 2, \ldots, N,$$

$$= p_{j-1}\frac{1}{t_\tau} \cdot t_\tau = p_{j-1}$$

In words: The first element of the $1 \times N$ parametric matrix P is found by subtracting from the Lagrangian function L, the product of the $n \times 1$ velocity matrix x_t by the $1 \times n$ Lagrangian momentum matrix p; and the remaining $N - 1 = n$ elements of the parametric momentum matrix are those of the

Lagrangian momentum matrix

Exercise 1 Show that the elements of the parametric momentum matrix are positively homogeneous functions of degree zero, of the $N \times 1$ matrix X_τ. Hint: Differentiate the relation $F(X, kX_\tau) = kF(X, X_\tau)$ with respect to X_τ

Exercise 2. Show that $P X_\tau = F$. Hint: Differentiate the relation $F(X, kX_\tau) = kF(X, X_\tau)$ with respect to k and then set $k = 1$

Exercise 3 Show that, if $L = T - V$ where T is a homogeneous function of degree 2 of the $n \times 1$ velocity matrix x_t, and V does not involve x_t, then $P_1 = -(T + V)$ Hint: $p = T_{x_t}$, $px_t = 2T$.

Lectures on Applied Mathematics

The Calculus of Variations

Lecture 2

Extremal Curves; The Euler-Lagrange Equation

The Lagrangian function L of a problem of Type 1 of the calculus of variations is a function of the $N \times 1$ matrix X and of the $n \times 1$ matrix x_t. We introduce the $(N + n) \times 1 = (2n + 1) \times 1$ matrix z whose first N elements are those of X and whose last n elements are those of x_t, and we regard the elements of z as the coordinates of a point in a space of $2n + 1$ dimensions. This space, which is known as the state space of the mechanical system whose Lagrangian function is L, is, like the time-coordinate space of the system, a numerical rather than a metrical space. We endow this numerical state-space with a Euclidean metric and we consider a region, i.e., an open, connected $(2n + 1)$ - dimensional domain D, in this $(2n + 1)$ - dimensional state-space over which $L = L(z)$ is, by hypothesis, a continuously differentiable function of z. If $X = X(t)$, $a \leqslant t \leqslant b$, is any smooth, or piecewise-smooth, curve C in the time-coordinate space, each point of C at which C is smooth furnishes a point of the state-space, the first N elements of the corresponding matrix z being those of $X(t)$ and the last n elements of z being those of x_t. The various

points of the state-space which we obtain in this way constitute a curve Γ which we term the image of C in the state-space. We shall confine our attention to those piecewise-smooth curves C in the time-coordinate space whose images Γ in the state-space are covered by the region D over which $L = L(z)$ is, by hypothesis, a continuously differentiable function of z.

When our problem of Type 1 of the calculus of variations is presented parametrically, our integral I appears as

$$I = \int_{\alpha}^{\beta} F(X, X_{\tau})\, d\tau$$

the curve C in the N-dimensional time-coordinate space along which I is evaluated being furnished by equations of the form

$$X = X(\tau), \quad \alpha \leqslant \tau \leqslant \beta .$$

The point z of our (2n + 1) - dimensional state-space, which is furnished by any point of C at which C is smooth, has as its first N coordinates the N elements of $X(\tau)$ and as its last n coordinates the n mutual ratios of the N elements of $X_{\tau}(\tau)$, it being assumed that $X_{\tau}(\tau)$ is not the zero $N \times 1$ matrix save, possibly, at a finite number of points of the interval $\alpha \leqslant \tau \leqslant \beta$. Thus z is independent of the particular parameter chosen to describe the curve C of integration. The parametric integrand $F(X, X_{\tau})$ is a positively homogeneous function of degree 1 of the $N \times 1$ matrix X_{τ}, and the value of the

integral I is independent of the parameter adopted to describe C.

In order to gain some idea as to how the integral I varies when the curve C along which it is evaluated is varied, we consider the following 1 - parameter family of piecewise-smooth curves

$$C_s:\ X(T, s) = X(T) + s\xi(T);\ \alpha \leq T \leq \beta;\ -\delta \leq s \leq \delta$$

Here s is the parameter which names the various members of the 1 - parameter family, and we suppose that s varies over a closed interval which is centered at $s = 0$. We observe that C_s reduces to C when $s = 0$, and we express this fact by the statement that we have imbedded C in the 1 - parameter family of curves C_s. $\xi(T)$ is any convenient N x 1 matrix which is piecewise-smooth over $\alpha < T < \beta$ and which reduces to the zero N x 1 matrix when $T = \alpha$ and when $T = \beta$, so that all the curves of our 1 - parameter family have the same end-points. We suppose that the various images Γ_s, in the state-space, of the curves C_s of our 1 - parameter family are all covered by the region D of our state-space over which L is, by hypothesis, a continuously differentiable function of z. Since

$$F_X = t_T L_X;\ P_1 = L - px_t,\ P_j = p_{j-1},\ j = 2, \ldots, N$$

it follows that F is, over D, a continuously differentiable function of the 2N x 1 matrix $Z = \begin{pmatrix} X \\ X_T \end{pmatrix}$ whose first N elements are those of X and whose last N elements are those of X_T. When I is

evaluated along any curve C_s of our 1 - parameter family, its value is a function of s which is furnished by the formula

$$I(s) = \int_{\alpha}^{\beta} F\left(X(\tau) + s\xi(\tau),\ X_{\tau}(\tau) + s\xi_{\tau}(\tau)\right) d\tau .$$

$I(s)$ is, for each value of s in the interval $-\delta < s < \delta$, a differentiable function of s, its derivative being

$$I_s(s) = \int_{\alpha}^{\beta} F_X\left(X(\tau) + s\xi(\tau),\ X_{\tau}(\tau) + s\xi_{\tau}(\tau)\right)\xi(\tau) d\tau$$

$$+ \int_{\alpha}^{\beta} F\left(X(\tau) + s\xi(\tau),\ X_{\tau}(\tau) + s\xi_{\tau}(\tau)\right) \xi_{\tau}(\tau) d\tau$$

We denote by δI the differential of $I(s)$ at $s = 0$ and we term δI the variation of I. Thus δI is the product of $I_s(0)$ by ds, where ds is an arbitrary number (which we may take to be 1). Similarly, we denote by δX the differential, with respect to s, of $X(\tau, s)$ at $s = 0$, and by δX_{τ} the differential, with respect to s, of $X_{\tau}(\tau, s)$ at $s = 0$ so that

$$\delta X = ds.\xi(\tau); \quad \delta X_{\tau} = ds.\xi_{\tau}(\tau)$$

We observe that

$$(\delta X)_{\tau} = \delta X_{\tau}$$

and we express this result by the statement that the order of variation and differentiation with respect to τ is, when these operations are applied to $X(\tau, s)$, immaterial. In this notation,

then, we have

$$\delta I = \int_a^B \left\{ F_X(X(\tau), X_\tau(\tau))\, \delta X + P(X(\tau), X_\tau(\tau))\, \delta X_\tau \right\} d\tau$$

In order to simplify this expression we observe that

$F_X(X(\tau), X_\tau(\tau))$ is continuous over $a \leqslant \tau \leqslant B$ save, possibly, for a finite number of points, namely, the points which furnish the points of C at which x_τ is not defined. Thus the 1 x N matrix function

$$G(\tau) = \int_a^\tau F_X(X(\tau), X_\tau(\tau))\, d\tau$$

is defined and is continuous at all points of the interval $a \leqslant \tau \leqslant B$. If $\tau = \tau_1$ is a point of the interval $a \leqslant \tau \leqslant B$ at which x_t is not defined, both $G(\tau_1 - o)$ and $G(\tau_1 + o)$ exist and are equal, their common value being $G(\tau_1)$. Furthermore, at those points of C at which x_τ is defined, $G_\tau = F_X$ so that $dG = d\tau . G_\tau = d\tau . F_X$. Thus the matrix product $F_X(X(\tau), X_\tau(\tau))\, \delta X$ may be integrated by parts to yield

$$\int_a^B \left\{ F_X(X(\tau), X_\tau(\tau))\, \delta X \right\} d\tau = (G\delta X)\Big|_a^B - \int_a^B \left\{ G(\delta X)_\tau \right\} d\tau$$

and, since $\delta X = ds . f(\tau)$ is zero, by hypothesis, at $\tau = a$ and at $\tau = B$, this reduces to $-\int_a^B \left\{ G(\delta X)_\tau \right\} d\tau$. Hence, since

$\delta X_\tau = (\delta X)_\tau$, δI appears as

$$\delta I = \int_a^B \left\{ (P - G)(\delta X)_\tau \right\} d\tau$$

It is clear that δI will be zero for all allowable choices of $\xi(\tau)$ if the $1 \times N$ matrix $P - G$ is constant along C; indeed, if this is the case, $\int_\alpha^\beta \left\{(P - G)(\delta X)_\tau\right\} d\tau$ is of the form

$$\sum_{j=1}^{N} A_j \int_\alpha^\beta (\delta X^j)_\tau \, d\tau = \sum_{j=1}^{N} A_j \, \delta X^j \Big|_\alpha^\beta ,$$ where A_j, $j=1,\ldots, N$, is constant, and this is zero since $\delta X^j = \xi^j(\tau)\, ds$, $j = 1, \ldots, N$, is zero, by hypothesis, at $\tau = \alpha$ and at $\tau = \beta$. This sufficient condition for δI to be zero, for all allowable choices of $\xi(\tau)$, is also necessary. To see this we observe that δI may be written in the form $\int_\alpha^\beta \left\{(P - G - A)(\delta X)_\tau\right\} d\tau$, where A is any constant $1 \times N$ matrix, and we consider the $1 \times N$ matrix function, $H(\tau) = \int_\alpha^\tau (P - G - A)\, d\tau$. It is clear that $H(\alpha) = 0$, and we may determine A by means of the formula $(\beta - \alpha)A = \int_\alpha^\beta (P - G) d\tau$ so that $H(\beta) = 0$. $H(\tau)$ is piecewise-smooth over $\alpha \leqslant \tau \leqslant \beta$, its derivative at any point τ of this interval, which furnishes a point of C at which C is smooth, being $P - G - A$. Hence we may take as our $N \times 1$ matrix $\xi(\tau)$ the transpose $H^*(\tau)$ of $H(\tau)$, and, when we do this, $(\delta X)_\tau = ds.(P - G - A)^*$ at all the points of C at which C is smooth. Thus

$\delta I = ds \int_\alpha^\beta \left\{(P - G - A)(P - G - A)^*\right\} d\tau$ is the product by ds of

the integral along C of the squared magnitude of the $1 \times N$ matrix $P - G - A$, and so for δI to be zero, $P - G - A$ must be zero at all the points of the interval $\alpha \leqslant \tau \leqslant \beta$ at which it is continuous. Thus we have the following important result:

The necessary and sufficient condition that δI, when evaluated along C, be zero for all allowable choices of the $N \times 1$ matrix $\xi(\tau)$, is that $P = G + A$ at all the points of C at which C is smooth, the $1 \times N$ matrix A being constant along C.

Since $G(\tau)$ is differentiable, with derivative F_X, at all the points of the interval $\alpha \leqslant \tau \leqslant \beta$ which furnish smooth points of C it follows that at all smooth points of C,

$$P_\tau = F_X$$

This equation is known as the Euler-Lagrange equation for problems of Type 1 of the calculus of variations. We term any piecewise-smooth curve along which it holds an extremal curve of F or of the Lagrangian function L. The laws of mechanics for systems which possess a potential energy function may be stated as follows:

The paths or curves in the time-coordinate space, traced by the mechanical system are extremal curves of the Lagrangian function L, or, equivalently, of the parametric integrand F.

In differentiating the relation $P = G + A$, in order to obtain the Euler-Lagrange equation $P_\tau = F_X$, we have lost the fact that A is

constant along C, since the relation $P_\tau = F_X$ would hold if A were merely piecewise-constant along C. If τ_1 is a point of the interval $\alpha \leqslant \tau \leqslant \beta$ which furnishes a point of C at which x_t is not defined, we know that $G(\tau_1 - 0) = G(\tau_1 + 0)$ and this implies that $P(\tau_1 - 0) = P(\tau_1 + 0)$. The parametric momentum matrix $P(\tau)$ is not defined at τ_1 but, on assigning to it at τ_1 the common value of $P(\tau_1 - 0)$ and $P(\tau_1 + 0)$, we see that it is defined and continuous at all the points of the extremal curve C. This is a remarkable fact since the velocity matrix x_t is not defined at the points of C at which C fails to be smooth. The last n coordinates of P are those of the Lagrangian momentum matrix $p = L_{x_t}$ which we may regard as a function of x_t, the $N \times 1$ matrix X being held fixed. Let us denote by $L_{x_t^*}$ the $n \times 1$ matrix which is the transpose p^* of the Lagrangian momentum matrix p and let us suppose that $L_{x_t^*}$ is, over the region D of our $(2n+1)$ - dimensional state-space, a continuously differentiable function of the $n \times 1$ velocity matrix x_t. Then the Jacobian matrix $L_{x_t^* x_t}$ of $p^* = L_{x_t^*}$ with respect to x_t is a symmetric $n \times n$ matrix of which the element in the j th row and k th column is $L_{x_t^j x_t^k}$, $j, k = 1, \ldots, n$. If this matrix is

nonsingular over **D**, the relation $p = L_{x_t}$ defines, over **D**, x_t as a function of X and p, and so the relation $p(\tau_1 - 0) = p(\tau_1 + 0)$, which is a consequence of the relation $P(\tau_1 - 0) = P(\tau_1 + 0)$, forces the equality $x_t(\tau_1 - 0) = x_t(\tau_1 + 0)$. In words:

Any extremal curve C of L is, when the n-dimensional matrix $L_{x_t^* x_t}$ exists and is continuous and nonsingular over **D**, not merely piecewise-smooth but actually smooth, x_t existing and being continuous at all the points of C.

If we assume, in addition, that the $n \times 1$ matrix $L_{x_t^*}$ is, over **D**, a continuously differentiable function of the $(2n+1) \times 1$ matrix $z = \begin{pmatrix} X \\ x_t \end{pmatrix}$ and not merely of the $n \times 1$ matrix x_t, then the theory of implicit functions assures us that the function $x_t(X, p)$ of X and p which is defined implicitly by the formula $L_{x_t^*} = p^*$ is a continuously differentiable function of the $(2n+1) \times 1$ matrix $\begin{pmatrix} X \\ p^* \end{pmatrix}$. Since this $(2n+1) \times 1$ matrix is continuously differentiable along C, it follows that x_t is continuously differentiable along C so that x_{tt} exists and is continuous at all the points of C. In words:

Any extremal curve C of L is, when the $n \times (2n+1)$ matrix $L_{x_t^* z}$ exists and is continuous over D (its

n-dimensional sub-matrix $L_{x_t^* x_t}$ being nonsingular over D) not merely smooth but possessed of continuous curvature, x_{tt} being defined and continuous at all the points of C.

Along C the parametric momentum matrix P is continuously differentiable, its derivative being furnished by the Euler-Lagrange equation, $P_{\tau} = F_X$. The last n of the equations furnished by this $1 \times N$ matrix equation may be written as $p_t = L_x$ or, equivalently, on taking the transpose of this $1 \times n$ matrix equation, as

$$L_{x_t^* x_t} x_{tt} + L_{x_t^* X} X_t = L_{x^*}$$

This $n \times 1$ matrix equation, which furnishes x_{tt}, along any extremal curve of L, as a function of the $(2n + 1) \times 1$ matrix $z = \begin{pmatrix} X \\ x_t \end{pmatrix}$, is the Euler-Lagrange equation for extremals which possess continuous curvature. We shall from now on suppose that L possesses the following two properties which guarantee that all its extremals possess continuous curvature:

1) $L_{x_t^* z}$ exists and is continuous over D.

2) $L_{x_t^* x_t}$ is nonsingular over D.

Exercise 1. Show that if F or, equivalently, L does not involve any given one of the elements X^j say, of X then the corresponding element P_j of the parametric momentum matrix is constant along any extremal curve C of F.

Exercise 2. Show that the Principle of Conservation of Energy holds for any mechanical system whose Lagrangian function does not involve t explicitly. Hint: $P_1 = -(T + V)$. Note: In view of the result of this exercise, a mechanical system whose Lagrangian function does not involve t explicitly is termed conservative.

The Calculus of Variations

Lecture 3

Lagrangian Functions Which Are Linear in x_t

When defining an extremal curve C of a given Lagrangian function L we imbedded C, which we assumed at the beginning to be merely piecewise-smooth, in a 1-parameter family of curves

$$\mathbf{X}(\tau, s) = \mathbf{X}(\tau) + s\xi(\tau), \quad \alpha \leqslant \tau \leqslant \beta, \quad -\delta \leqslant s \leqslant \delta$$

where $\xi(\tau)$ is any convenient $N \times 1$ matrix which is piecewise-smooth over $\alpha \leqslant \tau \leqslant \beta$ and which vanishes at $\tau = \alpha$ and at $\tau = \beta$. We found out later that if L satisfies some not very restrictive conditions, C must, if it is to qualify as an extremal curve of L, be smooth and, in addition, possess continuous curvature. Despite this fact, it is convenient to permit the comparison curves $\mathbf{X} = \mathbf{X}(\tau, s)$, $\alpha \leqslant \tau \leqslant \beta$, $-\delta \leqslant s \leqslant \delta$, to be only piecewise-smooth so that $\xi(\tau)$ may fail to be differentiable at a finite number of points of the interval $\alpha \leqslant \tau \leqslant \beta$. We do not impose additional restrictions upon our extremal curve if we imbed it in a family, $\mathbf{X} = \mathbf{X}(\tau, s)$, $\alpha \leqslant \tau \leqslant \beta$, $-\delta \leqslant s \leqslant \delta$, which does not involve the parameter s linearly, provided that the $N \times 1$ matrix function $\mathbf{X}(\tau, s)$ is such that $\delta\mathbf{X}_\tau = (\delta\mathbf{X})_\tau$ or, equivalently, that the two

mixed second-order derivatives $X_{\tau s}$ and $X_{s \tau}$ exist and are equal save, possibly for a finite number of points of the interval $\alpha \leqslant \tau \leqslant \beta$. This will certainly be the case if these derivatives exist and are continuous over the rectangle $\alpha \leqslant \tau \leqslant \beta$, $-\delta \leqslant s \leqslant \delta$, with the possible exception of a finite number of values of τ. We may also imbed our extremal curve C in a k - parameter family where k is any integer > 1. In this case s is a $k \times 1$ matrix, the parameter matrix of the family, and $X_{\tau s}$, $X_{s \tau}$ are $N \times k$ matrices which we assume to exist and be continuous over the $(k + 1)$ - dimensional interval $\alpha \leqslant \tau \leqslant \beta$ $-\delta^j \leqslant s^j \leqslant \delta^j$, $j = 1, \ldots, k$. This assumption guarantees the equality of these two $N \times k$ matrices, the possible existence of a finite number of exceptional values of τ being always allowed for. We assume that when s is the $k \times 1$ zero matrix, the curve $X = X(\tau, s)$, $\alpha \leqslant \tau \leqslant \beta$, of our k - parameter family reduces to the curve C in which we are interested, so that $X(\tau, 0) = X(\tau)$, $\alpha \leqslant \tau \leqslant \beta$. X_s is now a $N \times k$ matrix and the variation δX of X is defined by the formula

$$\delta X = X_s(\tau, 0)\, ds$$

where ds is an arbitrary $k \times 1$ matrix. Our formula for δI has precisely the same form as before; namely,

$$\delta I = \int_{\alpha}^{\beta} \left\{ F_X \left(X(\tau), X_\tau(\tau) \right) \delta X + P \left(X(\tau), X_\tau(\tau) \right) \delta X_\tau \right\} d\tau$$

and the right-hand side of this equation is of the form

$\delta_1 I + \dots + \delta_k I$ where $\delta_1 I$, for example, is furnished by the formula

$$\delta_1 I = \left[\int_{\alpha}^{\beta} \left\{ F_X (X(\tau), X_\tau(\tau)) f_1(\tau) + P (X(\tau), X_\tau(\tau)) (f_1(\tau))_\tau \right\} d\tau \right] ds^1$$

where $f_1(\tau) = X_{s^1}(\tau, 0)$. δI will be zero for an arbitrary choice of $ds = \begin{pmatrix} ds^1 \\ \vdots \\ ds^k \end{pmatrix}$ if, and only if, each of the partial variations $\delta_1 I, \dots, \delta_k I$ is zero, and each of these partial variations will be zero, by the argument given in Lecture 2, for arbitrary piecewise-smooth $N \times 1$ matrices $f_1(\tau), \dots, f_k(\tau)$, all of which vanish at $\tau = \alpha$ and at $\tau = \beta$ if, and only if, the Euler-Lagrange equation, $P_\tau = F_X$, holds along C.

Amongst the extremal curves of L or of F are to be found, if any such exist, the minimal curves and the maximal curves of L. A piecewise-smooth curve C in the N - dimensional time-coordinate space is said to be a minimal curve of L if the inequality

$$I(0) \leq I(s)$$

is valid, no matter what is the family of piecewise-smooth curves in which C is imbedded, provided only that the parametric interval $-\delta < s < \delta$ (or $-\delta^j < s^j < \delta^j$, $j = 1, \dots, k$, if the family is a k - parameter one) is sufficiently small. It follows from the theory

of maxima and minima of functions of one or more riables that a necessary, but by no means a sufficient, condition for C to be a minimal curve of L is that C be an extremal curve of L. For a maximal curve of L the inequality $I(0) \leqslant I(s)$ is reversed to read $I(0) \geqslant I(s)$, and it follows that a maximal curve of L is a minimal curve of $-L$. There is, then, no loss of generality in confining our attention to minimal curves, and we shall do this. Since every minimal curve of L, if any such exists, is an extremal curve of L we know that if L possesses the two properties described in Lecture 2; namely,

1) $L_{x_t^* z}$ exists and is continuous over D,

2) $L_{x_t^* x_t}$ is nonsingular over D,

then every minimal curve of L which is covered by D is not only smooth but, in addition, possesses continuous curvature. The question of the existence of minimal curves of L is a difficult one, compared to which the question of the existence of extremal curves of L is relatively superficial. For mechanical systems L is always a quadratic function, not necessarily homogeneous, of the elements of x_t, the coefficients of this quadratic function depending, in general on the $N \times 1$ time-coordinate matrix X. We continue, however, to use the symbol L to stand for the nonparametric integrand of any problem of Type 1 of the calculus of variations whether or not

this problem is one connected with the motion of a mechanical system, and we proceed to discuss a problem where L has a particularly simple form:

<u>L is a linear function of x_t</u>

Writing $L = rx_t + q$, where r is a $1 \times n$ matrix, we assume that the $1 \times N$ matrix $\mathbf{R} = (q, r)$ is a continuously differentiable function of the $N \times 1$ matrix $\mathbf{X} = \begin{pmatrix} t \\ x \end{pmatrix}$ over some region of the N - dimensional time-coordinate space which covers the curves C in which we are interested. Here $\mathbf{F} = \mathbf{R}\mathbf{X}_\tau$ so that $\mathbf{P} = \mathbf{R}$. Taking the transpose of this $1 \times N$ matrix equation, we have $\mathbf{P}^* = \mathbf{R}^*$ so that $\mathbf{P}^*_\tau = \mathbf{R}^*_{\mathbf{X}}\mathbf{X}_\tau$, where $\mathbf{R}^*_{\mathbf{X}}$ is the N - dimensional matrix of which the element in the j th row and kth column is $(\mathbf{R}_j)_{\mathbf{X}^k}$, $j, k = 1, \ldots, N$. Writing $\mathbf{F}$ in the form $\mathbf{X}^*_\tau \mathbf{R}^*$ we have $\mathbf{F}_{\mathbf{X}} = \mathbf{X}^*_\tau \mathbf{R}^*_{\mathbf{X}}$ or, equivalently, $\mathbf{F}_{\mathbf{X}^*} = \left(\mathbf{R}^*_{\mathbf{X}}\right)^* \mathbf{X}_\tau$, and so the Euler-Lagrange equation of our problem of Type 1 of the calculus of variations appears in the form

$$S\,\mathbf{X}_\tau = 0$$

where S is the alternating N - dimensional matrix $\mathbf{R}^*_{\mathbf{X}} - \left(\mathbf{R}^*_{\mathbf{X}}\right)^*$. If S is nonsingular. $\mathbf{X}_\tau$ must be the zero $N \times 1$ matrix and so our problem does not, in general, possess extremals. Observe that $L_{x^*_t} = r^*$ so that $L_{x_t^* x_t}$ is the zero n-dimensional matrix; if,

then, L does possess an extremal, this extremal need not possess continuous curvature nor even be smooth.

If the matrix S is identically zero over a region of our N-dimensional time-coordinate space which covers the curve C in which we are interested, the Euler-Lagrange equation $SX_\tau = 0$ becomes vacuous so that any piecewise-smooth curve which is covered by this region is an extremal curve of L. The element in the j th row and the k th column of S, j, k = 1, . . . , N, is $(R_j)_{x^k} - (R_k)_{x^j}$ and so S is identically zero when, and only when, **R** is the gradient of a differentiable point function $\phi(\mathbf{X})$ over a region of our time-coordinate space. When this is the case,

$$F = (\mathrm{grad}\,\phi)\,\mathbf{X}_\tau = \phi_\tau \quad \text{and } I = \int_\alpha^\beta \phi_\tau \, d\tau = \phi(\mathbf{X}(\beta)) - \phi(\mathbf{X}(\alpha))$$

are the same for all the curves of the family in which we suppose C to be imbedded. Thus $I(S) - I(0) = 0$ so that not only is every piecewise-smooth curve which is covered by the region of our time-coordinate space, over which we suppose S to be the zero N-dimensional matrix, an extremal curve of L but every such curve is at once a minimal and a maximal curve of L.

Exercise

Show that any two Lagrangian functions, L_1 and L_2, whose difference is of the form $rx_t + q$, where $R = (q, r)$ is the gradient of a differentiable point function over a region of the

N-dimensional time-coordinate space, possess the same extremal curves over this region. Note: It follows from the result of this exercise that if we are only interested in the extremal curves of L, we may neglect in L all terms of the type $rx_t + q$ where $R = (q, r) = \operatorname{grad} \phi(X)$. In particular, we may neglect all terms of the type $rx_t + q$, where the $1 \times n$ matrix r and the point function q are constants.

Lecture 4

The Legendre Condition for a Minimal Curve

We assume that our Lagrangian function $L = L(z)$ possesses, over a given region **D** of the $(2n+1)$ - dimensional state-space, the following properties.

1) L is, over **D**, a continuously differentiable function of the $(2n+1) \times 1$ matrix $z = \begin{pmatrix} X \\ x_t \end{pmatrix}$.

2) The $n \times (2n+1)$ matrix $L_{x_t^* z}$ exists and is continuous over **D** and the n - dimensional sub-matrix $L_{x_t^* x_t}$ of this $n \times (2n+1)$ matrix is nonsingular over **D**.

Then any piecewise-smooth curve C in the **N** - dimensional time-coordinate space whose image Γ in the $(2n+1)$ - dimensional state-space is covered by **D**, is an extremal curve of L if, and only if,

1) C is smooth and possesses continuous curvature;

2) C satisfies the Euler-Lagrange equation $p_t = L_x$ or, equivalently, $p^*_t = L_{z^*}$; i.e.,

$$L_{x^*_t x_t} x_{tt} + L_{x_t^* X} X_t = L_{x^*} .$$

This Euler-Lagrange equation defines x_{tt}, along C, as a continuous function of $z = \begin{pmatrix} X \\ x_t \end{pmatrix}$, and if this function of z is

not only continuous but continuously differentiable the existence and uniqueness theorem for ordinary differential equations assures us that there passes through each point of **D** an unambiguously determinate image Γ of an extremal curve C of L. We shall then, in order to avail ourselves of this uniqueness theorem, make the following additional assumption concerning the Lagrangian function L

3) $L_{x_t^*} = p^*$ possesses, over **D**, continuous second-order derivatives with respect to the elements of the $(2n+1) \times 1$ matrix z, and L_{x^*} is, over **D**, a continuously differentiable function of z.

We know, then, that there passes through each point of **D** an unambiguously determinate curve Γ which is the image of an extremal curve C of L. This curve C is not necessarily a minimal curve of L, and we proceed to investigate conditions which are imposed on L along C by the requirement that C be a minimal curve of L. This investigation of necessary conditions on L along C for C to be a minimal, and not merely an extremal, curve of L is a preliminary one; our ultimate goal is the procurement of conditions on L which are sufficient, i.e. strong enough, to ensure that the given extremal curve C of L will be actually a minimal curve of L. It may well happen that the

necessary conditions which we obtain may not be strong enough to be sufficient and, conversely, it may well happen that the sufficient conditions which we obtain may be too strong, in the sense that they are not necessary. The ideal would be a set of conditions on L which are at once necessary and sufficient to ensure that a given extremal curve C of L should be a minimal curve of L; but we shall not achieve this ideal in these lectures.

Note: The n equations implied by the $n \times 1$ matrix equation, $L_{x^*_t x_t} x_{tt} + L_{x^*_t x} x_t = L_{x^*}$, are the last n of the N equations implied by the $N \times 1$ matrix equation $P^*_\tau = F_{X^*}$, or, equivalently, $F_{X^*_\tau X_\tau} X_{\tau\tau} + F_{X^*_\tau X} X_\tau = F_{X^*}$, on the assumption that $t = t(\tau)$ possesses a continuous second derivative over $\alpha \leq \tau \leq \beta$. The question naturally arises: What about the first of these N equations? The fact that F is a positively homogeneous function, of degree 1, of X_τ assures us that $PX_\tau = F$ and, on differentiating this equation with respect to τ along C, we obtain

$P_\tau X_\tau + PX_{\tau\tau} = F_X X_\tau + PX_{\tau\tau}$, or, equivalently,

$(P_\tau - F_X) X_\tau = 0$. If, then, the last n of the N equations implied by the matrix equation $P_\tau = F_X$ are satisfied, so also will be the first (since $t_\tau > 0$ at all but a finite number

of points of the interval $\alpha \leqslant \tau \leqslant \beta$ and since P_τ and F_X are continuous over this interval).

The first necessary condition we shall determine which an extremal curve C of L must satisfy in order that it may qualify as a possible minimal curve of L concerns the n - dimensional matrix $L_{x^*_t x_t}$, which we have already assumed to be continuously differentiable and nonsingular over D. This condition, known as the Legendre condition, states that:

The quadratic form $u^* L_{x^*_t x_t} u$, where u is any real $n \times 1$ matrix of unit magnitude, must not be negative at any point of C.

This condition is what we may term a local condition; it concerns the nature of the n - dimensional matrix $L_{x^*_t x_t}$ at a given (but arbitrarily chosen) point of C. To prepare for the proof of the Legendre condition, which proof we shall complete in the next lecture, let t_1 be any interior point of the interval $a \leqslant t \leqslant b$ and let the equations of C, in terms of the Lagrangian parameter t, be

$$X = X(t), \quad a \leqslant t \leqslant b$$

Since we are concerned only with the behavior of $L_{x^*_t x_t}$ at the point $X(t_1)$ of C, we consider a positive number δ' which is so small that the closed interval $t_1 - \delta' \leqslant t \leqslant t_1 + \delta'$ is covered

by the interval $a \leqslant t \leqslant b$, and we denote by s any positive number $\leqslant \delta'$. Let $\mathfrak{f}(t)$ be the $n \times 1$ matrix function of t that is defined as follows:

$$\mathfrak{f}(t) = o \quad \text{if} \quad a \leqslant t \leqslant t_1 - s' \quad \text{and if} \quad t_1 + s' \leqslant t \leqslant b$$

$$\mathfrak{f}(t) = (t - t_1 + s')\, u \quad \text{if}, \quad t_1 - s' \leqslant t \leqslant t_1$$

$$\mathfrak{f}(t) = (t_1 + s' - t)\, u \quad \text{if} \quad t_1 \leqslant t \leqslant t_1 + s'$$

where u is any given constant $n \times 1$ matrix of unit magnitude, so that $u^* u = 1$. Then $\mathfrak{f}(t)$ fails to be differentiable at the three points $t_1 - s'$, t_1, $t_1 + s'$ of the interval $a \leqslant t \leqslant b$ and the various curves $C_{s,s'}$ of the 2 - parameter family

$$\overline{\mathbf{X}} = \mathbf{X}(t, s, s') = \mathbf{X}(t) + s\mathfrak{f}(t), \quad a \leqslant t \leqslant b, \quad -\delta \leqslant s \leqslant \delta, \quad -\delta' \leqslant s' \leqslant \delta',$$

are piecewise-smooth over $a \leqslant t \leqslant b$. The curves of this family for which $s = o$ coincide, no matter what is the value of s' in the interval $-\delta' \leqslant s' \leqslant \delta'$, with our given extremal curve C and, when $s \neq o$, the curve $C_{s,s'}$ coincides with C save for the interval $t_1 - s' \leqslant t \leqslant t_1 + s'$, which interval we may take to be as small as we find convenient. All the images $\Gamma_{s,s'}$ of the curves $C_{s,s'}$ are covered by $\mathbf{D}$ if δ is sufficiently small, but we shall need in our argument a bounded and closed subset of $\mathbf{D}$ which covers all the images $\Gamma_{s,s'}$ of the curves $C_{s,s'}$ and, more than this, a sufficiently small neighborhood of each point of each of these images. The reason for this need is

that we shall have to avail ourselves of the fact that any function which is continuous over a bounded and closed point-set is bounded over this point-set; our region $\mathbf{D}$ is open and we have no assurance that a function which is continuous over $\mathbf{D}$ is bounded over $\mathbf{D}$.

In order to construct the bounded and closed subset of $\mathbf{D}$ which we require, we first observe that there is no lack of generality in taking $\mathbf{D}$ to be bounded; for the image, Γ of $\mathbf{C}$ is bounded and we may replace $\mathbf{D}$, if it is unbounded, by the intersection of $\mathbf{D}$ and any open $(2n + 1)$ - dimensional interval which covers Γ. The points of the various images $\Gamma_{s,s'}$ of the curves $\mathbf{C}_{s,s'}$ of our 2 - parameter family constitute a closed subset of $\mathbf{D}$, since the three intervals $a \leqslant t \leqslant b$, $-\delta \leqslant s \leqslant \delta$, $-\delta' \leqslant s' \leqslant \delta'$ are all closed, and we denote by d any positive number which is less than one-half the distance of this closed subset of $\mathbf{D}$ from the closed boundary $\mathbf{D}'$ of $\mathbf{D}$. Finally, we denote by $\mathbf{D}_d$ the closed subset of $\mathbf{D}$, which consists of those points of $\mathbf{D}$ whose distance from $\mathbf{D}'$ is $\geq d$. Since the distance of any point of any one of the images $\Gamma_{s,s'}$ from $\mathbf{D}'$ is $> 2d$, it is clear that all the images $\Gamma_{s,s'}$ are covered by $\mathbf{D}_d$: the distance of any point of our $(2n + 1)$ - dimensional space, whose distance from any point of any one of the images $\Gamma_{s,s'}$ is $< d$, is by virtue of the triangle inequality $> 2d - d = d$, and so $\mathbf{D}_d$ covers any d-neighborhood of any point of any one of the images $\Gamma_{s,s'}$.

If C is to qualify as a minimal curve of L, the difference $I(s, s') - I(0, s')$ must be nonnegative provided only that δ is sufficiently small, $I(0, s')$ being, no matter what the value of s', the integral of L along C. This difference is the integral of $L(\bar{X}, \bar{x}_t) - L(X, x_t)$ over the interval $t_1 - s' \le t_1 \le t_1 + s'$, and we write $L(\bar{X}, \bar{x}_t) - L(X, x_t)$ in the form $\Delta_1 + \Delta_2$, where

$$\Delta_1 = L(\bar{X}, \bar{x}_t) - L(X, \bar{x}_t)$$

$$\Delta_2 = L(X, \bar{x}_t) - L(X, x_t)$$

Thus

$$I(s, s') - I(0, s') = \int_{t_1 - s'}^{t_1 + s'} \Delta_1 \, dt + \int_{t_1 - s'}^{t_1 + s'} \Delta_2 \, dt$$

and we shall treat, in detail, in our next lecture the two integrals on the right-hand side of this equation.

The Calculus of Variations

Lecture 5

Proof of the Legendre Condition

We have seen that the difference $I(s, s') - I(0, s')$ between the integrals of $L = L(z)$ evaluated along the curves $C(s, s')$ and $C(0, s') = C$ is the sum of the two integrals

$$I_1 = \int_{t_1 - s'}^{t_1 + s'} \Delta_1 \, dt \; ; \quad \Delta_1 = L(\bar{X}, x_t) - L(X, \bar{x}_t)$$

$$I_2 = \int_{t_1 - s'}^{t_1 + s'} \Delta_2 \, dt \; ; \quad \Delta_2 = L(X, \bar{x}_t) - L(X, x_t)$$

and we now proceed to appraise, one after the other, these two integrals. We have endowed our $(2n + 1)$ - dimensional state-space with a Euclidean metric, and the distance between the two points $\begin{pmatrix} X \\ \bar{x}_t \end{pmatrix}$ and $\begin{pmatrix} \bar{X} \\ \bar{x}_t \end{pmatrix}$ of this space is $s\,|t - t_1 + s'|$, which is $\leqslant \delta\,\delta'$. If, then, δ is sufficiently small, the point $\begin{pmatrix} X \\ \bar{x}_t \end{pmatrix}$ of our state-space is, since $\bar{z} = \begin{pmatrix} \bar{X} \\ \bar{x}_t \end{pmatrix}$ is a point of the image $\Gamma_{s,s'}$ of $C_{s,s'}$, covered by the point-set D_d of this state-space as is, also, every point of the line segment which connects the two points $\begin{pmatrix} X \\ \bar{x}_t \end{pmatrix}$ and $\begin{pmatrix} \bar{X} \\ \bar{x}_t \end{pmatrix}$ of the state-space.

The Theorem of the Mean of differential calculus assures us that $\Delta_1 = L_{\mathbf{X}}(\mathbf{X}', \bar{x}_t)(\bar{\mathbf{X}} - \mathbf{X})$, where $z' = \begin{pmatrix} \mathbf{X}' \\ \bar{x}_t \end{pmatrix}$ is some point of the line segment which connects the two points $\begin{pmatrix} \mathbf{X} \\ \bar{x}_t \end{pmatrix}$ and $\begin{pmatrix} \bar{\mathbf{X}} \\ \bar{x}_t \end{pmatrix}$ and, since $L_{\mathbf{X}}$ is, by hypothesis, continuous over $\mathbf{D}$ it is bounded over the closed subset $\mathbf{D}_d$ of $\mathbf{D}$; in other words, there exists a positive constant $\mathbf{M}$ which is such that the magnitude of the $1 \times N$ matrix $L_{\mathbf{X}}(\mathbf{X}', \bar{x}_t)$ is $\leq \mathbf{M}$ for all values of t in the interval $t_1 - s' \leq t < t_1 + s'$. Since the magnitude of the $N \times 1$ matrix $\bar{\mathbf{X}} - \mathbf{X}$ is $\leq \delta\delta'$, it follows that the absolute value of Δ_1, the product of the second of these two matrices by the first, is $\leq \mathbf{M}\delta\delta'$ and so

$$I_1 \leq 2\,\mathbf{M}\,\delta\delta' s'$$

Turning now to the integral I_2, we observe that, by virtue of the Extended Theorem of the Mean of differential calculus,

$$\Delta_2 = p(\mathbf{X}, x_t)(\bar{x}_t - x_t) + \frac{1}{2}(\bar{x}_t - x_t)^{*} L_{x_t^{*} x_t}(\mathbf{X}, v)(\bar{x}_t - x_t)$$

where $\begin{pmatrix} \mathbf{X} \\ v \end{pmatrix}$ is some point of the line segment which connects the two points, $\begin{pmatrix} \mathbf{X} \\ x_t \end{pmatrix}$ and $\begin{pmatrix} \mathbf{X} \\ \bar{x}_t \end{pmatrix}$ of our state-space. The distance between these two points is δ and so, by the same argument as before, the point $\begin{pmatrix} \mathbf{X} \\ v \end{pmatrix}$ is covered by $\mathbf{D}_d$ if δ is sufficiently small. Over the interval $t_1 - s' \leq t \leq t_1$,

$\bar{x}_t - x_t$ is the constant $n \times 1$ matrix $s\,u$ while, over the interval $t_1 < t < t_1 + s'$, it is the constant $n \times 1$ matrix $-s\,u$. Hence, the integral of $p_1(\bar{x}_t - x_t)$, where p_1 denotes the value of $p(X, x_t)$ at $t = t_1$, over the interval $t_1 - s' \leqslant t \leqslant t_1 + s'$ is zero, so that

$$\int_{t_1-s'}^{t_1+s'} \left\{ p(X, x_t)(\bar{x}_t - x_t) \right\} dt = \int_{t_1-s'}^{t_1+s'} \left\{ \left[p(X, x_t) - p_1 \right] (\bar{x}_t - x_t) \right\} dt$$

By virtue of the continuity of the $1 \times n$ matrix p along C, we know that the magnitude of the $1 \times n$ matrix $p(X, x_t) - p_1$ is arbitrarily small, say $< \epsilon$, over the interval $t_1 - s' \leqslant t \leqslant t_1 + s'$ if s' is sufficiently small. The magnitude of the $n \times 1$ matrix $\bar{x}_t - x_t$ is $\leqslant \delta$ over this interval and so the absolute value of the matrix product $\left[p(X, x_t) - p_1 \right] (\bar{x}_t - x_t)$ is $\leqslant \epsilon\delta$ over the interval $t_1 - s' \leqslant t \leqslant t_1 + s'$. Hence the absolute value of the integral $\int_{t_1-s'}^{t_1+s'} \left\{ p(X, x_t)(\bar{x}_t - x_t) \right\} dt$ is $\leqslant 2\epsilon\delta s'$,

where ϵ is arbitrarily small if the positive number δ' is sufficiently small.

Finally, the distance between the two points $z = \begin{pmatrix} X \\ x_t \end{pmatrix}$ and $z_1 = \begin{pmatrix} X(t_1) \\ x_t(t_1) \end{pmatrix}$ of our state-space is arbitrarily small over the

interval $t_1 - s' \leq t \leq t_1 + s'$, if δ' is sufficiently small, by virtue of the continuity of z over the interval $a \leq t \leq b$. We denote by J_1 the n-dimensional matrix $L_{x^*_t x_t}$ evaluated at $t = t_1$, and we suppose that there exists some $n \times 1$ matrix u of unit magnitude which is such that $u^* J_1 u$ is negative. Denoting the value of $\frac{1}{2} u^* J_1 u$ by $-k$ so that $k > 0$ we can choose δ and δ' sufficiently small so that the absolute value of the difference between the two quadratic forms $\frac{1}{2} u^* L_{x^*_t x_t} (\mathbf{X}, v)\, u$ and $\frac{1}{2} u^* J_1 u$ is $< \frac{1}{3} k$: indeed, the distance between the two points $\begin{pmatrix} \mathbf{X}(t_1) \\ x_t(t_1) \end{pmatrix}$ and $\begin{pmatrix} \mathbf{X} \\ v \end{pmatrix}$ of our state-space is arbitrarily small if δ and δ' are sufficiently small, and the n-dimensional matrix $L_{x^*_t x_t}$ is, by hypothesis continuous over $\mathbf{D}$. Hence, $\frac{1}{2} u^* L_{x^*_t x_t} (\mathbf{X}, v)\, u$ is negative and $< -\frac{2}{3} k$ over the interval $t_1 - s' \leq t \leq t_1 + s'$ so that

$$\frac{1}{2} \int_{t_1 - s'}^{t_1 + s'} \left\{ u^* L_{x^*_t x_t} (\mathbf{X}, v)\, u \right\} dt \quad \text{is negative and} \quad < -\frac{2}{3} k\, s'.$$

Since $\bar{x}_t - x_t = \pm s\, u$ over the interval $t_1 - s' < t < t_1 + s'$, it follows that

$$\frac{1}{2} \int_{t_1 - s'}^{t_1 + s'} \left\{ (\bar{x}_t - x_t)\, L_{x^*_t x_t} (\mathbf{X}, v)\, (\bar{x}_t - x_t) \right\} dt$$

is negative and $< -\frac{2}{3} k\, s^2 s'$. Choosing δ' so small,

δ and s remaining fixed, so that $2\epsilon\delta < \frac{1}{3}k s^2$, it follows that I_2 is negative and $< -\frac{1}{3}k s^2 s'$. Now we have seen that the numerical magnitude of I_1 is $\leqslant 2M\delta\delta' s'$ and, choosing δ' so small, δ and s remaining fixed, so that $2M\delta\delta' < \frac{1}{6}k s^2$, we see that $I_1 + I_2$ is negative, being $< -\frac{1}{6}k s^2 s'$. Hence, C is not a minimal curve of L and we have completed the proof of the Legendre condition:

An extremal curve C of L is not a minimal curve of L if the quadratic form, $\frac{1}{2}u^* L_{x^*_t x_t} u$, evaluated at any point of C, is negative for any $n \times 1$ matrix u of unit magnitude.

Note: In our argument we have assumed that the point t_1 of the interval $a \leqslant t \leqslant b$, which furnished the point of C at which the quadratic form $\frac{1}{2}u^* L_{x^*_t x_t} u$ was evaluated, was an interior point of this interval. However, if this quadratic form is negative at one of the end-points of this interval, it is negative, by virtue of the continuity, along C of the n-dimensional matrix $L_{x^*_t x_t}$ at interior points of the interval which are sufficiently near this end point.

The n - dimensional matrix $L_{x^*_t x_t}$ is symmetric and so it may be transformed at any point of our extremal curve C by means of a rotation matrix into diagonal form, the transforming

rotation matrix R depending, in general, upon the particular point of C which we have selected. The diagonal elements of the diagonal form $R^* L_{x^*_t x_t} R$ of $L_{x^*_t x_t}$, which are all real, are the characteristic numbers, $\lambda_1, \ldots, \lambda_n$, of $L_{x^*_t x_t}$, and our quadratic form $\frac{1}{2} u^* L_{x^*_t x_t} u$ appears as

$$\frac{1}{2}\left\{\lambda_1 (v^1)^2 + \ldots \quad + \lambda_n (v^n)^2\right\}$$

where $v^1, \ldots, v^n$ are the elements of a $n \times 1$ matrix $v = R^* u$ of unit magnitude. Since the $n \times 1$ matrix u of unit magnitude is arbitrary, so also is the $n \times 1$ matrix v, u being $R v$, where v is an arbitrary $n \times 1$ matrix of unit magnitude. It follows that, in order that our extremal curve C of L may qualify as a possible minimal curve of L, none of the characteristic numbers, $\lambda_1, \ldots, \lambda_n$, of L can be negative at any point of C. Indeed, if one of them, λ_j say, $j = 1, \ldots, n$, is negative at some point of C, we need only set all the elements of v, save v^j, zero, so that $v^j = \pm 1$, to obtain

$$\frac{1}{2} u^* L_{x^*_t x_t} u = \frac{1}{2} v^* R^* L_{x^*_t x_t} R v = \frac{1}{2} \lambda_j, \text{ which is negative.}$$

Conversely if all the characteristic numbers, $\lambda_1, \ldots, \lambda_n$, of $L_{x^*_t x_t}$ are nonnegative at every point of C,

$\frac{1}{2} u^* L_{x^*_t x_t} u = \frac{1}{2} \left\{ \lambda_1 (v^1)^2 + \dots + \lambda_n (v^n)^2 \right\}$ is nonnegative at every point of C and C qualifies as a possible minimal curve of L.

The determinant of the n - dimensional matrix $L_{x^*_t x_t}$ is the product, $\lambda_1 \dots \lambda_n$, of its characteristic numbers and so, if $L_{x^*_t x_t}$ is nonsingular along C (which assumption guarantees the smoothness of C), none of the characteristic numbers of $L_{x^*_t x_t}$ is zero at any point of C. In order, then, that C may qualify as a possible minimal curve of L, all the characteristic numbers of $L_{x^*_t x_t}$ must be positive (and not merely nonnegative) at every point of C. The statement that all the characteristic numbers of $L_{x^*_t x_t}$ are positive is equivalent to (that is, implies and is implied by) the statement that the quadratic form $\frac{1}{2} u^* L_{x^*_t x_t} u$ is positive no matter what is the n x 1 matrix u of unit magnitude; when this is the case we say that the n - dimensional matrix $L_{x^*_t x_t}$ is positively definite. When $L_{x^*_t x_t}$ is positively definite along C, we say that L satisfies, along C, the strong Legendre condition; as opposed to this, we say that L satisfies, along C, the weak Legendre condition when we are merely assured that all the characteristic numbers of $L_{x^*_t x_t}$

are nonnegative at every point of C. In other words, to pass from the weak to the strong Legendre condition, we have to add the requirement that the n - dimensional matrix $L_{x^*_t x_t}$ be non-singular at every point of C.

The characteristic numbers of $L_{x^*_t x_t}$ are the zeros of the polynomial function of λ, of degree n, $\det\left(\lambda E_n - L_{x^*_t x_t}\right)$, where E_n is the n - dimensional identity matrix. Fortunately, however, we do not have to determine these zeros in order to decide whether or not $L_{x^*_t x_t}$ is positively definite. Indeed,

$$\det\left(\lambda E_n - L_{x^*_t x_t}\right) = \lambda^n - I_1 \lambda^{n-1} + I_2 \lambda^{n-2} - \ldots + (-1)^n I_n$$

where I_j, $j = 1, \ldots, n$, is the sum of the various principal minors, of dimension j, of the n - dimensional matrix $L_{x^*_t x_t}$. For example, I_1 is the sum of the diagonal elements, i.e. the trace, of $L_{x^*_t x_t}$ and I_n is the determinant of $L_{x^*_t x_t}$. In general, I_j is the sum of the products (j at a time) of the n characteristic numbers $\lambda_1, \ldots, \lambda_n$ of $L_{x^*_t x_t}$. Therefore, if $L_{x^*_t x_t}$ is positively definite so that all its characteristic numbers are positive, all the n numbers, $I_1, \ldots, I_n$, are positive.

Conversely, if all these n numbers are positive, each term of the polynomial function, $\lambda^n - I_1 \lambda^{n-1} + \ldots + (-1)^n I_n$, has, when λ is negative, the same sign, so that this polynomial function has no negative zeros.

Thus

L satisfies along C the strong Legendre condition if, and only if, all the n numbers, $I_1, \ldots, I_n$, are positive at each point of C.

<u>Example</u>. For a mechanical system with n degrees of freedom, L is of the form $T(X, x_t) - V(X)$, where T is a quadratic polynomial function, not necessarily homogeneous, of the $n \times 1$ velocity matrix x_t. Writing the terms of T which are of the second degree in the elements of x_t in the form $\frac{1}{2} x^*_t G x_t$, where $G = G(X)$ is a n - dimensional matrix point function in our time-coordinate space, our n - dimensional matrix $L_{x^*_t x_t}$ is simply G; the transpose p^* of the $1 \times n$ momentum matrix p is $G x_t$. Thus L satisfies the weak Legendre condition along C if, and only if, all the characteristic numbers of G are nonnegative at each point of C, in which case we say that G is positively semi-definite along C. On the other hand, L satisfies the strong Legendre condition along C if, and only if, G is positively definite along C. If $n = 1$, so that our mechanical system

has but one degree of freedom, G is a 1 - dimensional matrix, i.e., a point function, in the (t, x) - plane, and, in order that an extremal curve C of L may qualify as a possible minimal curve of L, this point function must be nonnegative at every point of C.

Lectures on Applied Mathematics

The Calculus of Variations

Lecture 6

Constrained Problems; The Hamilton Canonical Equations

We now suppose that the curves in our N - dimensional time-coordinate space along which the integral $I = \int_a^b L\,dt$ of a problem of Type 1 of the calculus of variations is evaluated; are all constrained to lie on an n-dimensional point locus. We term a problem, of Type 1, of the calculus of variations, where such a supposition is made, a constrained problem. We shall suppose that the equation of the n - dimensional point locus in our N - dimensional time-coordinate space is of the form $f(X) = 0$, where the function f is continuously differentiable over some convenient connected region R of the time-coordinate space and we term the equation $f(X) = 0$ the constraint imposed on our problem. A simple example of such a constrained problem is the following:

Given two points on the sphere $t^2 + (x^1)^2 + (x^2)^2 = 1$ in 3 - dimensional Euclidean space, does there exist on this sphere a rectifiable curve joining these two points whose length is least, and, if so, is this curve unambiguously determinate?

The argument for a constrained problem proceeds as in the case of an ordinary, or unconstrained, problem the only difference being that we must always respect, in the constrained problem, the constraint $f(\mathbf{X}) = 0$. Confining our attention to smooth extremal curves of L and writing I in the parametric form, $\int_{\alpha}^{\beta} \mathbf{F}\, d\tau$, we have

$$\delta I = \int_{\alpha}^{\beta} \left\{ \mathbf{F}_{\mathbf{X}}\, \delta \mathbf{X} + \mathbf{P}\, \delta \mathbf{X}_{\tau} \right\} d\tau$$

where the parametric momentum matrix $\mathbf{P} = \mathbf{F}_{\mathbf{X}_{\tau}}$ is, by virtue of our hypotheses concerning the Lagrangian function $L = L(\mathbf{X}, x_t)$, continuously differentiable along the curve C of integration so that the term $\mathbf{P}\, \delta \mathbf{X}_{\tau} = \mathbf{P}\, (\delta \mathbf{X})_{\tau}$ may be integrated by parts to yield, since $\delta \mathbf{X}$ is zero, by hypothesis, at $\tau = \alpha$ and at $\tau = \beta$, $\int_{\alpha}^{\beta} (\mathbf{P} \delta \mathbf{X}_{\tau})\, d\tau = -\int_{\alpha}^{\beta} (\mathbf{P}_{\tau}\, \delta \mathbf{X})\, d\tau$. Thus

$$\delta I = -\int_{\alpha}^{\beta} \left\{ \left(\mathbf{P}_{\tau} - \mathbf{F}_{\mathbf{X}} \right) \delta \mathbf{X} \right\} d\tau$$

We cannot, however, conclude that the necessary and sufficient condition that δI be zero for arbitrary allowable variations $\delta \mathbf{X}$ of the curve C of integration is that $\mathbf{P}_{\tau} = \mathbf{F}_{\mathbf{X}}$; this relation is obviously sufficient but it is too strong, i. e., it is not necessary. Indeed, these variations $\delta \mathbf{X}$ are subject to the relation $f_{\mathbf{X}}\, \delta \mathbf{X} = 0$,

since the family of comparison curves in which we suppose C to be imbedded all lie, by hypothesis, in the n - dimensional point locus $f(\mathbf{X}) = 0$. If, then, λ is an undetermined multiplier, i. e., any function of τ which is continuous over the interval $\alpha \leq \tau \leq \beta$, we may write $\delta \mathbf{I}$ in the form

$$\delta \mathbf{I} = - \int_{\alpha}^{\beta} \left\{ (\mathbf{P}_{\tau} - \mathbf{F}_{\mathbf{X}} - \lambda f_{\mathbf{X}}) \delta \mathbf{X} \right\} d\tau$$

and the relation $\mathbf{P}_{\tau} = \mathbf{F}_{\mathbf{X}} + \lambda f_{\mathbf{X}}$ is sufficient, no matter what the undetermined multiplier λ for the vanishing of $\delta \mathbf{I}$, for all allowable variations $\delta \mathbf{X}$ of the curve of integration. We proceed now to show that this condition, with an appropriate determination of λ, is also necessary for the vanishing of $\delta \mathbf{I}$ for all allowable variations $\delta \mathbf{X}$. We first observe that any point of our n - dimensional point locus $f(\mathbf{X}) = 0$, at which $f_{\mathbf{X}}$ is the zero $1 \times N$ matrix, is a singular point of this point locus and we assume that, if any such singular points exist, their number is finite. If, then, we can prove that $\mathbf{P}_{\tau} = \mathbf{F}_{\mathbf{X}} + \lambda f_{\mathbf{X}}$, with an appropriate determination of λ, at any point of C which is not a singular point of the point locus $f(\mathbf{X}) = 0$, it follows from the continuity along C of both sides of this equation that the equation continues to hold at those points of C, if any such exist, which are singular points of this point locus. Let, then, τ_1 be a point of the interval $\alpha < \tau < \beta$ at which one of the elements,

f_{X^j} say, of the 1 x N matrix f_X is not zero and let us set all the elements of the N x 1 matrix δX zero along C except δX^j and δX^k, where k is any one of the numbers, different from j, of the set 1, ..., N. We may assign arbitrary values to δX^k and then δX^j is determined at any point where $f_{X^j} \neq 0$, by means of the equation $f_{X^j}\,\delta X^j + f_{X^k}\,\delta X^k = 0$. We now determine the, as yet undetermined multiplier λ so that over an interval $\tau_1 - \delta \leq \tau \leq \tau_1 + \delta$ which is so small that f_{X^j} is different from zero over it (this being always possible since f_{X^j} is different from zero at τ_1 and continuous over $\alpha \leq \tau \leq \beta$) $(P_j)_\tau = F_{X^j} + \lambda f_{X^j}$; then δI reduces to

$$\delta I = -\int_{\tau_1-\delta'}^{\tau_1+\delta'} \left\{ \left[(P_k)_\tau - F_{X^k} - \lambda f_{X^k} \right] \delta X^k \right\} d\tau$$

on the understanding that δX^j and δX^k are set equal to zero at those points of the interval $\alpha \leq \tau \leq \beta$ which are not covered by the interval $\tau_1 - \delta' \leq \tau \leq \tau_1 + \delta'$. If $(P_k)_\tau - F_{X^k} - \lambda f_{X^k}$ were not zero at $\tau = \tau_1$ we could choose, by virtue of the continuity of this expression over the interval $\alpha \leq \tau \leq \beta$, δ' so small that $(P_k)_\tau - F_{X^k} - \lambda f_{X^k}$ is one-signed over the interval $\tau_1 - \delta' \leq \tau \leq \tau_1 + \delta'$. All we have to do to ensure that δI is not zero is to choose δX^k so that it also is one-signed over the interval $\tau_1 - \delta' \leq \tau \leq \tau_1 + \delta'$. Since k is any one of the numbers

different from j of the set $1, \ldots, N$ it follows that, at $\tau = \tau_1$ $P_\tau = F_X + \lambda f_X$, and, since τ_1 is any point of the interval $\alpha \leq \tau \leq \beta$ at which f_X is not the zero $1 \times N$ matrix, this implies that the equation

$$P_\tau = F_X + \lambda f_X$$

is valid at all the points of C. This equation is the Euler-Lagrange equation for constrained problems of Type 1 of the calculus of variations. Strictly speaking, λ is not determined at any point of C which is a singular point of the n - dimensional point locus $f(X) = 0$, but we define it at such points, if any such exist on C, by the requirement that λ be continuous along C.

Example. The constrained arc-length problem in N - dimensional Euclidean space.

Here $F = (X^*_\tau X_\tau)^{1/2}$, $P = \frac{X^*_\tau}{F}$, and the Euler-Lagrange equation is $\frac{X^*_{\tau\tau}}{F} - \frac{F_\tau X^*_\tau}{F^2} = \lambda f_X$. Choosing as our parameter the arc-length s along C, F has along C the constant value 1, and the Euler-Lagrange equation reduces to $X^*_{ss} = \lambda f_X$ or, equivalently, $X_{ss} = \lambda f_{X^*}$. When $f(X)$ is $\frac{1}{2}(X^*X - r^2)$, so that the point-locus whose equation is $f(X) = 0$ is a sphere of radius r with center at the origin, our Euler-Lagrange equation is $X_{ss} = \lambda X$. Thus along any extremal curve of this constrained

arc-length problem, $\mathbf{X} \cdot \mathbf{X}_{ss} = \lambda \mathbf{X} \cdot \mathbf{X} = \lambda r^2$. Since $\mathbf{X} \cdot \mathbf{X}_s = 0$ along any curve on the sphere whose equation is $\mathbf{X} \cdot \mathbf{X} = r^2$ and since $\mathbf{X}_s \cdot \mathbf{X}_s = 1$, $\mathbf{X} \cdot \mathbf{X}_{ss} + 1 = 0$ along our extremal curve so that our undetermined multiplier λ has the constant value $-\frac{1}{r^2}$ in this problem. The curvature vector $\mathbf{X}_{ss}$ has the direction opposite to that of the radius vector $\mathbf{X}$, and the magnitude of this curvature vector, being the product of the magnitude r of the radius vector by the absolute value of λ, is $\frac{1}{r}$. From the relation $\mathbf{X}_{ss} = \lambda \mathbf{X}$ we obtain $\mathbf{X}^j \mathbf{X}^k{}_{ss} - \mathbf{X}^k \mathbf{X}^j{}_{ss} = 0$, where j and k are any two different numbers from the set $1, \ldots, N$, and so $\mathbf{X}^j \mathbf{X}^k{}_s - \mathbf{X}^k \mathbf{X}^j{}_s$ has a constant value, a^{jk} say, along any extremal curve. This implies that if j, k, l is any triad of different numbers from the set $1, \ldots, N$, then $a^{jk} \mathbf{X}^l + a^{kl} \mathbf{X}^j + a^{lj} \mathbf{X}^k = 0$. In particular, we see that any extremal curve of the constrained arc-length problem on the sphere $(\mathbf{X}^1)^2 + (\mathbf{X}^2)^2 + (\mathbf{X}^3)^2 = r^2$, in 3-dimensional Euclidean space, is an arc of a great circle of this sphere obtained by intersecting it with a plane, $a^{23} \mathbf{X}^1 + a^{31} \mathbf{X}^2 + a^{12} \mathbf{X}^3 = 0$, which passes through the origin, i.e., the center of the sphere.

The $2n \times 1$ matrix $\begin{pmatrix} x \\ p \end{pmatrix}$ is commonly termed the phase matrix of a dynamical system, and we shall term the $2N \times 1$

matrix $\begin{pmatrix} X \\ p \end{pmatrix}$ the extended phase matrix of the system. Regarding the elements of the $2N \times 1$ matrix $\begin{pmatrix} X \\ p \end{pmatrix}$ as the coordinates of a point in a $2N$ - dimensional space, which we endow with a Euclidean metric, we refer to this space as the extended phase-space of the mechanical system or, generally, of any problem of Type 1, either unconstrained or constrained, of the calculus of variations. Now $P_1 = L(X, x_t) - p\, x_t$ and the right-hand side of this equation is, by virtue of our hypotheses concerning the Lagrangian function L, a continuously differentiable function of X and p. In the case of a conservative mechanical system, for which T is a homogeneous second-degree function $\frac{1}{2} x^*_t\, G\, x_t$ of the velocity $n \times 1$ matrix x_t whose matrix G is nonsingular, we have $p^* = G\, x_t$ so that $x_t = G^{-1} p^*$. Hence T may be expressed as $\frac{1}{2}\, p\, G^{-1} p^*$ and $P_1 = -\frac{1}{2}\, p\, G^{-1}\, p^* - V$. The right-hand side of this equation is the negative of $T + V$, when expressed as a function of X and p; we denote $T + V$, when expressed as a function of X and p, by $H(X, p)$ and we term $H(X, p)$ the Hamiltonian function of the conservative mechanical system. Similarly, for any problem of Type 1 of the calculus of variations, unconstrained or constrained, we denote $p\, x_t - L(X, x_t)$, when expressed in terms of X and p, by $H(X, p)$ and we term $H(X, p)$ the Hamiltonian function of the problem. Then, $P_1 + H(X, p) = 0$ so that the points $Z = \begin{pmatrix} X \\ p \end{pmatrix}$

of our extended phase-space are constrained to lie on the $(2N - 1)$ - dimensional point locus $\Phi(Z) = 0$, where $\Phi(Z) = P_1 + H(X, p)$.

Since our parametric integrand $F = F(X, X_\tau)$ is a positively homogeneous function, of degree 1, of the $N \times 1$ matrix X_τ, we have $F = F_{X_\tau} X_\tau = P X_\tau$, and so our integral $I = \int_\alpha^\beta F \, d\tau$ may be written in the form

$$I = \int P \, dX$$

We may regard the right-hand side of this equation as an integral extended along a curve $\bar{\Gamma}$ in our $2N$ - dimensional extended phase-space, $\bar{\Gamma}$ being constrained to lie in the $(2N - 1)$ - dimensional point locus $\Phi(Z) = 0$ in this space. The integrand $F = P X_\tau$ of this integral is a positively homogeneous function, of degree 1, of the $2N \times 1$ matrix $Z_\tau = \begin{pmatrix} X_\tau \\ P_\tau \end{pmatrix}$, and the corresponding parametric momentum matrix is the $1 \times 2N$ matrix $(P, 0)$. Since $F_Z = (0, X_\tau)$, the Euler-Lagrange equation of the corresponding constrained problem of Type 1 of the calculus of variations is

$$P_\tau = \lambda \Phi_X = \lambda H_X ; 0 = X_\tau + \lambda \Phi_P$$

The first of the last N of this set of $2N$ equations is $0 = t_\tau + \lambda$ and the remaining n of this set of N equations are equivalent to the matrix equation $x_\tau + \lambda H_p = 0$. Thus the

BLANK PAGE

Euler-Lagrange equation of our constrained calculus of variations problem may be written, since $P_1 = -H$, as follows

$$\frac{dx^j}{H_{p_j}} = \frac{dt}{1} = \frac{dp_k}{-H_{x^k}} = \frac{dH}{H_t} = -\lambda d\tau \ , \ j, k = 1, \ldots, n .$$

These equations are known as the Hamilton canonical equations of our unconstrained problem of Type 1 of the calculus of variations.

Lectures on Applied Mathematics

The Calculus of Variations

Lecture 7

The Reciprocity between L and H ; The Transversality Conditions

The Hamiltonian function $H(X, p)$ of an unconstrained problem of Type 1 of the calculus of variations is connected with the Lagrangian function $L(X, x_t)$ of this problem by means of the relation

$$L(X, x_t) + H(X, p) = p\, x_t$$

and the symmetry of this equation indicates that the relationship between the two functions L and H is a reciprocal one. On differentiating the relation just written with respect to x_t, X being held constant, we obtain $p + H_{p*}\, p^*_{x_t} = p + x^*_t\, p^*_{x_t}$, or, equivalently, $H_{p*}\, p^*_{x_t} = x^*_t\, p^*_{x_t}$ (Note: In differentiating the matrix product $p\, x_t$ with respect to x_t, we write this matrix product in the equivalent form $x^*_t\, p^*$ when differentiating the matrix factor p, in order to encounter the n - dimensional matrix $p^*_{x_t}$. p_{x_t} is not an n - dimensional matrix but, rather, a $1 \times n^2$ matrix,and the matrix product $p_{x_t}\, x_t$ does not exist.)

$p^*_{x_t}$ is the n - dimensional symmetric matrix $L_{x^*_t x_t}$ which we have assumed to be nonsingular over the region D of our (2n + 1) - dimensional state-space, and so the equation $H_{p^*} p^*_{x_t} = x^*_t p^*_{x_t}$ implies the equation $H_{p^*} = x^*_t$ or, equivalently, the equation

$$x_t = H_p$$

This equation plays, when we start with the Hamiltonian function $H = H(X, p)$ the role played by the equation $p = L_{x_t}$ when we start with the Lagrangian function $L = L(X, x_t)$. The n-dimensional matrix $H_{p\,p^*}$ is the Jacobian matrix of x_t with respect to p^*, so that it is the reciprocal of the n - dimensional matrix $L_{x^*_t x_t}$, which is the Jacobian matrix of p^* with respect to x_t. The N x 1 matrix X is held constant when calculating these Jacobian matrices. If, then, we start with the Hamiltonian function $H = H(X, p)$ and wish to determine the corresponding Lagrangian function $L = L(X, x_t)$, we assume that the n - dimensional matrix $H_{p\,p^*}$ exists and is continuous and nonsingular over a region of the (2n + 1) - dimensional $\begin{pmatrix} X \\ p \end{pmatrix}$ - space, so that the formula $x_t = H_p(X, p)$ defines p, over the corresponding region of the (2n +1) -

dimensional state-space, as a continuous function of the $(2n+1) \times 1$ matrix $z = \begin{pmatrix} X \\ x_t \end{pmatrix}$. Then $L(X, x_t)$ is furnished, over this region of the state-space, by the formula

$$L(X, x_t) = p\,x_t - H(X, p).$$

We see, then, that the n equations furnished by the matrix equation $x_t = H_p$ have nothing to do with the particular smooth curve $x = x(t)$, $a \leqslant t \leqslant b$ in our time-coordinate space. They merely restate in a different way the definition of the Lagrangian momentum matrix p, which was originally defined by means of the formula $p = L_{x_t}$. Any such smooth curve has an image Γ in the 2N - dimensional extended phase-space, since X and P are defined as functions of τ over the parametric interval $\alpha \leqslant \tau \leqslant \beta$, and Γ lies in the $(2N-1)$ - dimensional point locus whose equation is $\phi(Z) = 0$, where Z is the $2N \times 1$ matrix $\begin{pmatrix} X \\ P \end{pmatrix}$ and $\phi(Z) = P_1 + H(X, p)$. Along Γ, we have $x_t = H_p = \Phi_p$ and this equation may be extended, since $\Phi_{P_1} = 1$, to read $X_t = \Phi_P$. In other words, one-half of the 2N first-order differential equations of any extremal curve of our constrained problem in the 2N - dimensional extended phase-space are satisfied by the image, in this space, of any smooth curve in our N - dimensional time-coordinate space. We proceed to show that the

other half of our 2N equations are satisfied if, and only if, the curve in our N - dimensional time-coordinate space is an extremal curve of the Lagrangian function L of our unconstrained problem; in other words, they are equivalent to the Euler-Lagrange equation $\dot{P}_{\tau} = F_X$. Indeed, since $F = L\, t_{\tau}$, this Euler-Lagrange equation may be written in the equivalent form $P_t = L_X$ and on differentiating with respect to X, the relation $L(X, x_t) + H(X, p) = p\, x_t = x^*_t\, p^*$, x_t being held fixed, we obtain $L_X + H_X + H_{p^*} p^*_X = x^*_t\, p^*_X$. This implies, since $H_{p^*} = x^*_t$, that $L_X = - H_X$. Thus our Euler-Lagrange equation is equivalent to the equation $P_t = - H_X$, and this may be written, since $P_1 = - H(X, p)$, as $\frac{d\, p^j}{-H_{x_j}} = \frac{d\, H}{H_t} = d\, t$, $j = 1, \ldots, n$.

Thus we have the following important result:

The images, in the 2N - dimensional extended phase-space, of the extremal curves, in the N - dimensional time-coordinate space, of the Lagrangian function of our unconstrained calculus of variations problem are the extremal curves of the Lagrangian function $P\, X_{\tau}$ of the corresponding constrained calculus of variations problem in the 2N - dimensional extended phase-space.

Up to this point we have, when defining an extremal curve of a given Lagrangian function $L = L(z)$, required that the end-

points $A = X(\alpha)$ and $B = X(\beta)$ of the various members of our family of comparison curves should all be the same. We now investigate what form δI takes when these end-points vary, and, to indicate this variation, we write

$$A = A(s) = X(\alpha, s) \; ; \; B = B(s) = X(\beta, s)$$

It is often convenient to permit the parametric interval $\alpha \leqslant \mathcal{T} \leqslant \beta$ to vary with s so that α and β are functions, which we assume to be continuously differentiable, of the parameter, or parametric matrix, s. Then δI appears as

$$\delta I = F(\beta)\,\delta\beta - F(\alpha)\,\delta\alpha + \int_{\alpha}^{\beta} \left\{ F_X\,\delta X + P\,\delta X_{\mathcal{T}} \right\} d\mathcal{T}$$

where $F(\beta) = F(B, X_{\mathcal{T}}(\beta))$ and $F(\alpha) = F(A, X_{\mathcal{T}}(\alpha))$. If the curve of integration C is a smooth extremal curve of L or, equivalently, of F, the parametric momentum matrix P is continuously differentiable along C, and the integrand $P\,\delta X_{\mathcal{T}} = P(\delta X)_{\mathcal{T}}$ may be integrated by parts to yield

$$\int_{\alpha}^{\beta} \left\{ P\,\delta X_{\mathcal{T}} \right\} d\mathcal{T} = P(\beta)\,\delta X(\beta) - P(\alpha)\,\delta X(\alpha) - \int_{\alpha}^{\beta} \left\{ P_{\mathcal{T}}\,\delta X \right\} d\mathcal{T}$$

so that

$$\delta I = F(\beta)\,\delta\beta + P(\beta)\,\delta X(\beta) - \left\{ F(\alpha)\,\delta\alpha + P(\alpha)\,\delta X(\alpha) \right\} - \int_{\alpha}^{\beta} \left\{ (P_{\mathcal{T}} - F_X)\,\delta X \right\} d\mathcal{T}$$

Since the curve of integration C is, by hypothesis, a smooth extremal curve of L, $P_{\tau} - F_X$ is the zero $1 \times N$ matrix along it, so that our formula for δI reduces to

$$\delta I = \left\{ F(\beta)\,\delta\beta + P(\beta)\,\delta X(\beta) \right\} - \left\{ F(\alpha)\,\delta\alpha + P(\alpha)\,\delta X(\alpha) \right\}$$

This may be simplified further, since $F = F(X, X_{\tau})$ is a positively homogeneous function, of degree 1, of X_{τ}; thus $F(\beta) = P(\beta)\,X_{\tau}(\beta)$ and $F(\alpha) = P(\alpha)\,X_{\tau}(\alpha)$ and, since $\delta B = X_{\tau}(\beta)\,\delta\beta + \delta X(\beta)$ and $\delta A = X_{\tau}(\alpha)\,\delta\alpha + \delta X(\alpha)$, we have

$$\delta I = P(\beta)\,\delta B - P(\alpha)\,\delta A$$

<u>Warning:</u> You must avoid the error of identifying $\delta X(\alpha)$ with δA or $\delta X(\beta)$ with δB. A, for example, is the value of X when $\tau = \alpha$ so that it involves the parameter, or parametric matrix, s not only explicitly but also implicitly through $\alpha = \alpha(s)$; thus δA involves, in addition to the term $X_s(\alpha)\,d\,s = \delta X(\alpha)$, the term $X_{\tau}(\alpha)\,\delta\alpha$.

In order that the arc of our smooth extremal curve of L between the points A and B may qualify when these end-points of the arc may vary as a possible minimal curve of L, we require that δI be zero for all permissible variations of the end-points A and B. Holding A fixed so that $\delta A = 0$,

we obtain the necessary condition $P(\beta)\,\delta B = 0$; similarly, $P(\alpha)\,\delta A = 0$. The two conditions

$$P(\alpha)\,\delta A = 0\,; \quad P(\beta)\,\delta B = 0$$

which are necessary in order that our smooth extremal curve of L may qualify (when its end points may vary) as a possible minimal curve of L, are known as the Transversality Conditions.

Example. For the arc-length problem in N - dimensional space, $F = \left\{X^{*}_{\tau} X_{\tau}\right\}^{1/2}$, so that $P^{*} = \frac{X_{\tau}}{F}$ is the $N \times 1$ matrix, of unit magnitude, which furnishes the direction of the curve $X = X(\tau)$, $\alpha \leqslant \tau \leqslant \beta$, in the sense in which τ is increasing. Since F_X is the zero $1 \times N$ matrix, P or, equivalently, P^{*} is constant along any extremal curve of F and the transversality conditions express the fact that for any extremal curve to qualify (when its end points may vary) as a possible minimal curve of F, the extremal must be perpendicular to all possible directions in which either of the two end-points may move. For example, in the arc-length problem in the plane (if the end-points are required to lie one on each of two nonconcentric circles) any extremal curve, which is necessarily a line-segment, must pass (if it is to qualify as a possible minimal curve) when produced, if necessary,

through each of the two centers of the two given circles. This necessary condition is not sufficient; there are four line-segments on the line joining the two centers which have one end-point on one of the two circles and the other end-point on the other Of these four line-segments only one is a minimal curve, the other three being neither minimal nor maximal curves.

We now consider an n - dimensional locus of initial points A in our N - dimensional time-coordinate space so that the $N \times 1$ matrix A is a function, which we assume to be continuously differentiable, of an $n \times 1$ parametric matrix s. We suppose that we are given over this locus a continuously differentiable $n \times 1$ matrix function of s which we denote by $x_t(A)$, since we intend to consider the extremal curves of L which pass through the various points A and whose velocity $n \times 1$ matrices are furnished at A by the matrix function in question. This matrix function is assumed to be such that the points $z = \begin{pmatrix} A \\ x_t(A) \end{pmatrix}$ of our $(2N - 1)$ - dimensional state-space are all covered by the region D of this space over which we suppose that the $n \times (2N - 1)$ matrix $L_{x_t^* z}$ and the $1 \times N$ matrix L_X are continuously differentiable. As a result

through each point A of our n - dimensional locus an unambiguously determinate extremal curve of L, whose velocity matrix at A is furnished by the given $n \times 1$ matrix $x_t(A)$. We suppose further that the n - parameter family of extremal curves of L which we obtain in this way simply covers a region R of our N - dimensional time-coordinate space, where R is covered by the projection of D on this space. In other words, we suppose that there passes through each point X of R one, and only one, member of our n - parameter family.

The equations of our n - parameter family of extremals are of the form $X = X(\mathcal{T}, s)$, where the $N \times 1$ matrix X is at each point of R a continuously differentiable function of the $N \times 1$ matrix $S = \begin{pmatrix} \mathcal{T} \\ s \end{pmatrix}$. The statement that our n - parameter family of extremal curves of L simply covers R is equivalent to the statement that the N - dimensional Jacobian matrix X_S is nonsingular over R. Thus S is a continuously differentiable function of X over R and so x_t, being a continuously differentiable function of S, is a continuously differentiable function of X over R. We denote this $n \times 1$ matrix function of X by $w(X)$. Our $1 \times N$ parametric momentum matrix P is then a continuously differentiable function of X over R. When $n = 1$, so that our problem of Type 1 of the calculus of variations

is a plane problem, we term such a 1 - parameter family of extremal curves an extremal field over **R**, and we term the point function $w = w(t, x)$, the field function. When $n > 1$, however, it is necessary for our n - parameter family of extremal curves to possess a property, which is not required when $n = 1$, before we term it an extremal field over **R**. When our n - parameter family does possess this property, we refer to the $n \times 1$ matrix $w(X)$ as the field matrix. We shall discuss in our next lecture the property which our n - parameter family of extremals must possess, when $n > 1$, before it can qualify as an extremal field over **R**, and we shall begin to see the fundamental significance of extremal fields in the determination of sufficient conditions for an extremal curve of L to qualify as a minimal curve of L.

Lectures on Applied Mathematics

The Calculus of Variations

Lecture 8

Extremal Fields; The Hilbert Invariant Integral

We have seen that, if $I = \int_a^b L\,dt = \int_\alpha^\beta F\,d\tau$ is the integral of a problem of Type 1 of the calculus of variations evaluated along a smooth extremal curve C of L between two variable end-points A and B, then

$$\delta I = P(\beta)\,\delta B - P(\alpha)\,\delta A$$

We now suppose that C is a member of an n-parameter family of extremal curves of L of the type described in our last lecture (so that each member of the family intersects a given n - dimensional point locus which we regard as the locus of our initial points A, and the family covers simply a given region R of our N - dimensional time-coordinate space). A is, by hypothesis, a continuously differentiable function of an n x 1 parameter matrix s, and if X is any point of that member of our n - parameter family of extremal curves of L whose initial point is A, $X = X(\tau, s)$ is a continuously differentiable function of the N x 1 matrix $S = \begin{pmatrix} \tau \\ s \end{pmatrix}$. Setting τ equal to an arbitrarily chosen continuously differentiable function β of s, we may

take $B = X(\beta, s)$ as the second end-point of an extremal arc of L whose first end-point is A and avail ourselves of the relation $\delta I = P(\beta)\, \delta B - P(\alpha)\, \delta A$. Letting B vary along an arbitrary piecewise-smooth curve C_B^* connecting any two points B_1 and B_2 of R, A will vary along an unambiguously determinate piecewise-smooth curve C_A^*, in our given n dimensional point locus, connecting two points A_1 and A_2, of this locus which are unambiguously determined by the two arbitrarily given points B_1 and B_2, respectively, of R since the N-dimensional Jacobian matrix X_S is, by hypothesis, nonsingular over R. Denoting by I_1 and I_2 the values of I evaluated along the extremal curves C_1 and C_2 of L whose end-points are (A_1, B_1) and (A_2, B_2), respectively, we obtain, on integrating the relation $\delta I = P(\beta)\, \delta B - P(\alpha)\, \delta A$ with respect to the independent variable which identifies the points of the arbitrarily chosen piecewise-smooth curve C_B^*, the relation

$$I_2 - I_1 = \int_{C_B^*} P(\beta)\, \delta B \qquad \int_{C_A^*} P(\alpha)\, \delta A$$

This equation has the notable feature that its left-hand side is unambiguously determined by the two arbitrarily given points B_1 and B_2 of R and is, accordingly, insensitive to a change

of the particular piecewise-smooth curve C_B^* that connects these points. Thus the difference $\int_{C_B^*} P(\beta)\,\delta B - \int_{C_A^*} P(\alpha)\,\delta A$

depends only on the two points B_1 and B_2 and not at all on the particular piecewise-smooth curve C_B^* that connects these two points. When $n = 1$, our locus of initial points A in which the curve C_A^* must lie is 1 - dimensional, i. e. is a curve. Therefore C_A^* is unambiguously determined by the two points B_1 and B_2, and is, like the difference $\int_{C_B^*} P(\beta)\,\delta B - \int_{C_A^*} P(\alpha)\,\delta A$,

independent of the piecewise-smooth curve C_B^* which we choose to connect these two points. In this particular case, then, not only is this difference independent of C_B^* but so also is each of its two terms. Thus, when $n = 1$, so that our problem of Type 1 of the calculus of variations is a plane one, we have the following important result:

The integral $\int_{C_B^*} P(\beta)\,\delta B$ is independent of the particular

piecewise-smooth curve C_B^* that we select to connect the two arbitrarily given points B_1 and B_2 of R

When $n > 1$, this result is not, in general, true since the curve C_A^* in our n-dimensional locus of initial points is not unambiguously determined by its end points A_1 and A_2.

If, however, we so choose the $n \times 1$ velocity matrix $\dot{x}$, along our n - dimensional locus of initial points that the corresponding $1 \times N$ parametric momentum matrix $P(\alpha)$ is such that the integral $\int_{C_A^*} P(\alpha)\, \delta A$ is unambiguously determined by A_1 and A_2, it being the same for all piecewise-smooth curves in our n-dimensional locus of initial points which connect A_1 and A_2, then we are assured that $\int_{C_B^*} P(\beta)\, \delta B$ is independent of C_B^*.

Writing $P(\alpha)\, \delta A$ in the form $P(\alpha)\, A_s\, ds$ and assuming that our n - dimensional locus of initial points is simply connected, we see that a sufficient condition for $\int_{C_A^*} P(\alpha)\, \delta A$ to be independent of C_A^* is that $P(\alpha)\, A_s$ be the gradient, with respect to the $n \times 1$ matrix s, of a continuously differentiable function of s. In particular, if this function is a numerical constant so that $P(\alpha)\, A_s$ is the zero $1 \times n$ matrix, the integral $\int_{C_A^*} P(\alpha)\, \delta A$ is independent of C_A^*. It is, in fact, zero for all piecewise-smooth curves connecting the two points A_1 and A_2. In this case we say that our n - parameter family of extremal curves of L intersects transversally the n - dimensional locus of initial points; for example, in the arc-length problem, every member of our n - parameter family of

extremals, each a segment of a straight line, intersects at right angles our n - dimensional locus of initial points when the family intersects this point locus transversally.

We term our n - parameter family of extremal curves of L an extremal field over R when, and only when, the integral $\int_{C_B^*} P(B)\,\delta B$, which we shall denote by I^*, is independent of the particular piecewise-smooth curve C_B^* which we choose to connect the two arbitrarily given points B_1 and B_2 of R. At any point X of R, the velocity n x 1 matrix x_t is a continuously differentiable point function which we denote by w (X) and term the field matrix of the extremal field. When n = 1, any 1 - parameter family of extremal curves of L is an extremal field over R if it possesses the following two properties:

1) Each member of the family intersects a given piecewise-smooth curve in R.

2) R is simply covered by the family; in other words, there passes through each point of R one, and only one, member of the family. This will be the case if the 2 - dimensional matrix X_S is nonsingular over R.

When n > 1, our n - parameter family of extremal curves of L must possess, in addition to these two properties, the following third property before it can qualify as an extremal field over R:

3) $\int_{C_A^*} P(\alpha)\,\delta A$ must be independent of C_A^*: in other words, $P(\alpha)\,A_s\,ds$ must be an exact differential over our n - dimensional locus of initial points, which point locus we assume to be simply connected.

When our n - parameter family of extremal curves of L is a field over R we term the integral

$$I^* = \int_{C_B^*} P(\beta)\,\delta B$$

the Hilbert invariant integral of the extremal field. In this integral, $P(\beta) = P(B, w(B))$ where $w(X)$ is the $n \times 1$ field matrix of the extremal field. Note: $P = P(X, X_\tau)$ is a positively homogeneous function, of degree zero, of X_τ, so that

$$P(X, X_\tau) = P\left(X, \frac{X_\tau}{t_\tau}\right),$$ where $\dfrac{X_\tau}{t_\tau}$ is the $N \times 1$ matrix $\begin{pmatrix} 1 \\ x_t \end{pmatrix}$. Hence, $P(X, X_\tau)$ is a function of X and x_t, and no confusion is caused by writing this function as $P(X, x_t)$, since x_t is a $n \times 1$ matrix whereas X_τ is a $N \times 1$ matrix.

If the two points B_1 and B_2 happen to lie on a member of our extremal field and if the particular curve C_B^* connecting these two points which we select is an arc of this member, then

$P(\beta) = P(B, w(B)) = P(B, B_\mathcal{T})$, and I^* reduces to

$$\int_{C_B^*} P(B, B_\mathcal{T})\, B_\mathcal{T}\, d\mathcal{T} = \int_{\mathcal{T}_1}^{\mathcal{T}_2} F\, d\mathcal{T} .$$

Thus:

When the curve C_B^* along which we evaluate Hilbert's invariant integral I^* is an arc of a member of our extremal field, I^* is equal to the integral

$I = \int_{\mathcal{T}_1}^{\mathcal{T}_2} F\, d\mathcal{T}$ evaluated along this arc.

We have seen that our n - parameter family of extremal curves of L will be an extremal field of L over R if, in addition to intersecting a given n - dimensional point locus and simply covering R, it intersects this n - dimensional point locus transversally. By writing the equation of our n-dimensional point locus in the form $J(X) = 0$, we see that our n-parameter family of extremal curves of L will intersect this n-dimensional point locus transversally if, and only if, $\lambda P = J_X$ over it, since $J_X\, \delta X = 0$ is the only constraint on δX over the point locus. Here λ is an undetermined multiplier, and we ask what condition is imposed upon our n-dimensional locus of initial points by the requirement that λ have the constant value 1 over it.

Since, under this requirement, $J_t = P_1$, $J_x = p$, we see that $J_t + H(X, J_x) = 0$ over the n - dimensional locus of initial points. This first order partial differential equation, which we write as $\phi(X, J_X) = 0$, where $\phi(X, P) \equiv P_1 + H(X, p)$, is known as the Hamilton-Jacobi partial differential equation. Since it is insensitive to the addition of a constant to the point function $J = J(X)$, we see that if we set $P = J_X$ at each point of any member of the 1 - parameter family of n - dimensional point loci $J(X) = C$, where C is any constant, then the corresponding n - parameter family of extremal curves of L will cut transversally the n - dimensional point locus whose equation is $J(X) = C$. On differentiating the relation $\phi(X, J_X) = 0$ with respect to X, we obtain $\phi_X + \phi_{P^*} J_{X^*X} = 0$, where $J = J(X)$ is a solution, which we assume to possess continuous second derivatives with respect to X, of the Hamilton-Jacobi partial differential equation $\phi(X, J_X) \equiv J_t + H(X, J_x) = 0$.

Since $X_t = \phi_P$, by virtue of the defining relation $P = F_{X_T}$, we have

$$\phi_{X^*} = - J_{X^*X} X_t = - \left(J_{X^*}\right)_t$$

and it follows from the Euler-Lagrange equation $P_t = L_X = -\phi_X$

that $P^* \quad J_{X^*}$ is constant along any extremal curve of L. If, then, we set $P = J_X$ over our n - dimensional locus of initial points, this relation remains valid over R and, hence, over that part of the n - dimensional locus $J(X) = C$ which is covered by R. Thus:

If $J = J(X)$ is a solution possessing over R continuous second derivatives with respect to the $N \times 1$ matrix X and such that $J(X) = o$ is the equation of an n - dimensional point locus, then the n - parameter family of extremal curves of L determined by setting $P = J_X$ over this n dimensional point locus cuts transversally each member of the 1 - parameter family of point loci determined by the equation $J(X) = C$, where C is any constant; the relation $P = J_X$ being valid over each member of the family.

<u>Example</u>:

For the arc-length problem in N - dimensional space,

$F = \left(X^*_T \ X_T\right)^{1/2}$, $P = \dfrac{X^*_T}{F}$ so that $P\,P^* = 1$. For this problem, $L = \left(1 + x^*_t\,x_t\right)^{1/2}$, $p = \dfrac{x^*_t}{L}$,

$P_1 = \left(1 + x^*_t\,x_t\right)^{1/2} - \dfrac{x^*_t x_t}{L} = \dfrac{1}{L}$.

Now $L^2 = 1 + L^2 p p^*$, so that $\frac{1}{L} = (1 - p p^*)^{1/2}$ and so

$H(X, p) = -(1 - p p^*)^{1/2}$. The relation

$\phi(X, p) \equiv P_1 + H(X, p) = 0$ is equivalent to the relation

$P_1 = (1 - p p^*)^{1/2}$, and this is, in turn, equivalent to the relation

$(P_1)^2 = 1 - p p^*$, i.e. $P P^* = 1$. The Hamilton-Jacobi equation

is $J_t + H(X, J_X) = 0$, and this may be written in the more

symmetrical form $J_X J_{X^*} = 1$. Thus, when $n = 1$, the

Hamilton-Jacobi equation is $\left|J_t\right|^2 + \left|J_x\right|^2 = 1$. The solutions

of this equation which are linear functions of t and x are

$J(t, x) = \ell t + m x$, where $\ell^2 + m^2 = 1$; the 1 - parameter

family of extremals obtained by setting $P = J_X$ over the curve

$\ell t + m x = 0$ is the family of straight lines which intersect at

right angles the straight line whose equation is $\ell t + m x = 0$.

This 1 - parameter family of extremal curves of L also inter-

sects at right angles every member of the 1 - parameter family

$\ell t + n x = C$, where C is any constant. If we introduce plane polar coordinates (r, θ) by setting $t = r \cos \theta$, $x = r \sin \theta$, the equation $(J_t)^2 + (J_x)^2 = 1$ takes the form $J_r^2 + \frac{1}{r^2} J_\theta^2 = 1$, and the solutions of this equation which are functions of r alone are of the form r = constant. The 1 - parameter family of extremal curves obtained by setting $P = J_x$ over the circle $r = 1$ all intersect this circle orthogonally, and they also intersect orthogonally each member of the family of concentric circles $r = C$, where C is any positive constant.

Lectures on Applied Mathematics

The Calculus of Variations

Lecture 9

The Weierstrass E-Function; Positively Regular Problems

The importance of the Hilbert invariant integral

$$I^* = \int P(X, w(X))\, dX$$

in the problem of determining sufficient conditions for an extremal curve of the Lagrangian function L of a problem of Type 1 of the calculus of variations to be a minimal curve of L stems from the following two properties of I^*:

1) I^* is invariant; i.e., if C and $\bar{C}$ are any two piecewise-smooth curves connecting any two points A and B of the region R, of our N - dimensional time-coordinate space, then $I^*(\bar{C}) = I^*(C)$

2) I^* reduces to I when the curve of integration is a member of the extremal field over R whose $n \times 1$ matrix field function is $w(X)$.

To see how useful these two properties of I are, let C be an arc between two points A and B of R of a member of the extremal field of L over R whose matrix field function is

$w(X)$, and let $\bar{C}$ be any piecewise-smooth curve in R whose end-points are A and B. Then, in order that C may be a minimal curve of L, $I(\bar{C})$ must be $\geqslant I(C)$ if the domain D of our $(2n+1)$ - dimensional state-space, whose projection on the time-coordinate space covers R, is suitably restricted. Thus it is necessary to appraise $I(\bar{C}) - I(C)$, and this appraisal is difficult because the curves of integration C and $\bar{C}$ are different; we avoid this difficulty by first replacing I(C) by $I^*(C)$, which is legitimate because of Property 2, and then replacing $I^*(C)$ by $I^*(\bar{C})$, which is legitimate because of Property 1. Thus the difference $I(\bar{C}) - I(C)$ is the same as the difference $I(\bar{C}) - I^*(\bar{C})$, and this latter difference is relatively simple to appraise, since the curve of integration is the same in each of the two integrals. $\bar{C}$ is furnished by an equation of the form

$$\bar{X} = \bar{X}(\tau); \qquad \alpha \leqslant \tau \leqslant \beta$$

and $I^*(\bar{C}) = \int_\alpha^\beta \left\{ P(\bar{X}, w(\bar{X}))\, \bar{X}_\tau \right\} d\tau$. On denoting by $W(\bar{X})$ the $N \times 1$ matrix whose first element is $\bar{t}_\tau$ and whose remaining n elements are those of $\bar{t}_\tau w(\bar{X})$, we have $P(\bar{X}, w(\bar{X})) = P(\bar{X}, W(\bar{X}))$, where $P(\bar{X}, W(\bar{X})$ is a positively homogeneous function of degree zero, of the $N \times 1$ matrix $W(\bar{X})$.

Thus $I^*(\bar{C}) = \int_{\alpha}^{\beta} \left\{ P(\bar{X}, W(\bar{X}))\, \bar{X}_{\tau} \right\} d\tau$ and, since

$I(\bar{C}) = \int_{\alpha}^{\beta} F(\bar{X}, \bar{X}_{\tau})\, d\tau$, we have

$$I(\bar{C}) - I(C) = \int_{\alpha}^{\beta} \left\{ F(\bar{X}, \bar{X}_{\tau}) - \left[P(\bar{X}, W(\bar{X}))\, \bar{X}_{\tau} \right] \right\} d\tau$$

Since $P(\bar{X}, W(\bar{X}))\, W(\bar{X}) = F(\bar{X}, W(\bar{X}))$, we may write this equation in the form

$$I(\bar{C}) - I(C) = \int_{\alpha}^{\beta} E(\bar{X}, \bar{X}_{\tau}, W(\bar{X}))\, d\tau$$

where

$$E(\bar{X}, \bar{X}_{\tau}, W(\bar{X})) = F(\bar{X}, \bar{X}_{\tau}) - F(\bar{X}, W(\bar{X})) - P(\bar{X}, W(\bar{X}))(\bar{X}_{\tau} - W(\bar{X}))$$

If $X = X(\tau)$, $\alpha \leqslant \tau \leqslant \beta$, is any piecewise-smooth curve in R, the function

$$\begin{aligned} E(X, X_{\tau}, W(X)) &= F(X, X_{\tau}) - P(X, W(X))\, X_{\tau} \\ &= F(X, X_{\tau}) - F(X, W(X)) - P(X, W(X))(X_{\tau} - W(X)) \end{aligned}$$

is piecewise continuous along the curve,and $I(\bar{C}) - I(C)$ is the integral of this function along $\bar{C}$. $E(X, X_{\tau}, W(X))$ is a positively homogeneous function, of degree 1, of the two $N \times 1$ matrices X_{τ} and $W(X)$, i.e., if k is any positive number

$$E(X, k X_{\tau}, k W(X)) = k E(X, X_{\tau}, W(X)).$$

This function is known as the Weierstrass E - function, and our task is to appraise the integral of this Weierstrass E - function along $\bar{C}$. If E is nonnegative along $\bar{C}$, it is certain that $I(\bar{C}) > I(C)$; furthermore, this weak inequality may be replaced by the strong inequality $I(\bar{C}) > I(C)$ if E, in addition to being nonnegative along $\bar{C}$, is positive at a single one of its points of continuity on $\bar{C}$. Hence, C is a minimal curve of L if E is nonnegative over R, and, furthermore, if E is positive at all except, possibly, a finite number of points of any piecewise-smooth curve $\bar{C}$ in R connecting the end-points A and B of C, $I(C) < I(\bar{C})$, so that C furnishes an absolute minimum of our integral I. The importance of this result is that it furnishes a uniqueness theorem for minimal curves of L through the points A and B of R. Indeed, if C and $\bar{C}$ were two different minimal curves of L connecting the points A and B of R, we would have the two contradictory inequalities $I(C) < I(\bar{C})$, $I(\bar{C}) < I(C)$. This uniqueness theorem is on a plane of mathematical difficulty quite different from the more elementary uniqueness theorem of ordinary differential equations, which states that, if L satisfies over D certain not very restrictive conditions, then there passes through each point $z = \begin{pmatrix} X \\ x_t \end{pmatrix}$ of D an unambiguously determinate curve Γ

which is the image of an extremal curve of L We may term this a local uniqueness theorem, since the data which are necessary to determine the unique extremal curve of L are the coordinates of a single point z of D On the other hand, the uniqueness theorem which we are now encountering and which depends on the positiveness of the Weierstrass E - function, may be termed a global uniqueness theorem, since the data which are necessary to determine the unique minimal curve of L are the coordinates of two arbitrarily chosen points A and B of R In order to avail ourselves of the proof just given of the validity over R of this global existence theorem, we must be certain that our Lagrangian function L possesses the following two properties:

1) Any extremal curve of L that connects any two points A and B of R may be imbedded in an extremal field of L over R.

2) The Weierstrass E functions of the various extremal fields of L referred to in 1), are all essentially positive. By this we mean that these E - functions are nonnegative and such that the integral of E over any curve connecting the points A and B, other than the extremal curve which is imbedded in the field, is positive. For example, E is essentially positive if it is non-negative and is zero at not more than a finite number of points of

any piecewise-smooth curve connecting A and B other than the extremal curve of L referred to.

The arc-length problem in N - dimensional Euclidean space is one for which it is easy to prove the global existence theorem without the necessity of integrating the Euler-Lagrange equation, since the Weierstrass E - function takes a particularly simple form in this problem. We have already seen that any region of our N - dimensional time-coordinate space may be covered by an extremal field of $F = (X^{*}_{\tau} X_{\tau})^{1/2}$ consisting of parallel open line-segmerts. If A and B are any two points of this N - dimensional Euclidean space,the line-segment $A \longrightarrow B$ is an extremal curve of the Lagrangian function $L = (1 + x^{*}_{t} x_{t})^{1/2}$

or, equivalently, of the parametric integrand $F = (X^{*}_{\tau} X_{\tau})^{1/2}$,

and we take as our extremal field of L the field whose line-segments are parallel to the line-segment $A \longrightarrow B$. Then $P(\bar{X}, W(\bar{X}))$ is the unit $N \times 1$ matrix which furnishes the direction of the line-segment $A \longrightarrow B$ so that $P(\bar{X}, W(\bar{X}))\, \bar{X}_{\tau} = F(\bar{X}, \bar{X}_{\tau}) \cos\theta$, where θ is the angle from the line-segment $A \to B$ to the tangent, in the sense in which τ is increasing, to the curve $\bar{C}$, whose equation is $\bar{X} = \bar{X}(\tau)$, $a \leqslant \tau \leqslant B$. Hence

$E(\bar{X}, \bar{X}_{\tau}, W(\bar{X})) = F(\bar{X}, \bar{X}_{\tau})(1 - \cos\theta)$ which is nonnegative,

since $F(\bar{X}, \bar{X}_\tau) \geqslant 0$. Furthermore, $F(\bar{X}, \bar{X}_\tau) > 0$ except possibly, at a finite number of points of $\bar{C}$, since $\bar{C}$ is, by hypothesis, piecewise smooth, and so E is positive along $\bar{C}$ except for this finite number of points and the points where $\theta = 0$. If, then, there is one point of $\bar{C}$ not belonging to the finite number of points referred to, at which $\theta \neq 0$, the integral of E along $\bar{C}$ is positive so that $I(\bar{C}) > I(C)$. If $\theta = 0$, except possibly at a finite number of points of $\bar{C}$, $\bar{C}$ is a line-segment. and so must coincide with C, since it passes through the points A and B. This completes the proof of the global uniqueness theorem for the arc-length problem in N - dimensional Euclidean space.

It follows from the extended theorem of the mean of differential calculus that

$$
\begin{aligned}
E(\bar{X}, \bar{X}_\tau, W(\bar{X})) &\equiv F(\bar{X}, \bar{X}_\tau) - F(\bar{X}, W(\bar{X})) - P(\bar{X}, W(\bar{X}))(\bar{X}_\tau - W(\bar{X})) \\
&= \frac{1}{2}\left\{\bar{X}_\tau - W(\bar{X})\right\}^* F_{X^*_\tau X_\tau}(\bar{X}, V)(\bar{X}_\tau - W(\bar{X}))
\end{aligned}
$$

where $V = W(\bar{X}) + \theta\left\{\bar{X}_\tau - W(\bar{X})\right\}$, $0 < \theta < 1$. Since the first elements of the two $N \times 1$ matrices $\bar{X}_\tau$ and $W(\bar{X})$ are the same, namely $\bar{t}_\tau$, and since the n - dimensional matrix obtained by erasing the first row and the first column of the N - dimensional matrix $F_{X^*_\tau X_\tau}$ is the quotient of $L_{x^*_t x_t}$ by t_τ,

F being $L\, t_\tau$, $E(\bar{X}, \bar{X}_\tau, W(\bar{X}))$ may be written in the form

$\frac{1}{2}\left\{\bar{x}_t - w(\bar{X})\right\}^* L_{x_t^* x_t}(\bar{X}, v)\left\{\bar{x}_t - w(\bar{X})\right\}$ times t_τ , where $w(X)$ is the $n \times 1$ matrix field function, and $v = w(\bar{X}) + \theta\left\{\bar{x}_t - w(\bar{X})\right\}$.

Thus

$$I(\bar{C}) - I(C) = \frac{1}{2}\int_a^b \left\{\bar{x}_t - w(\bar{X})\right\}^* L_{x_t^* x_t}(\bar{X}, v)\left\{\bar{x}_t - w(\bar{X})\right\} dt$$

We say that our problem of Type 1 of the calculus of variations is positively regular over our $(2n + 1)$ - dimensional domain D, if the n - dimensional matrix $L_{x_t^* x_t}$ is positively definite over D. We assume that D has the type of convexity which is implied by the statement that if $\begin{pmatrix} X \\ w_1 \end{pmatrix}$ and $\begin{pmatrix} X \\ w_2 \end{pmatrix}$ are any two points of D which have the same projection on R, then the line segment connecting these two points is covered by D. Since $\begin{pmatrix} \bar{X} \\ w(\bar{X}) \end{pmatrix}$ and $\begin{pmatrix} \bar{X} \\ \bar{x}_t \end{pmatrix}$ are, by hypothesis, points of D, so also is $\begin{pmatrix} \bar{X} \\ v \end{pmatrix}$ and so, for a positively regular problem, E is positive at those points of $\bar{C}$ for which $\bar{x}_t - w(\bar{X})$ is not the zero $n \times 1$ matrix. Hence, $I(\bar{C}) > I(C)$ unless $\bar{x}_t = w(\bar{X})$ along $\bar{C}$; when this is the case, $\bar{C}$ is a member of our extremal field over R and, since it passes through A, it must coincide with C. Consequently, we see that C furnishes an absolute minimum of $I(\bar{C})$ for all

piecewise-smooth curves $\bar{C}$ connecting the points A and B which are covered by R. Thus the global uniqueness theorem is valid for positively regular problems of Type 1 of the calculus of variations (provided that L possesses the Property 1 concerning the possibility of imbedding any extremal curve of L, covered by R, in an extremal field of L over R).

Lectures on Applied Mathematics

The Calculus of Variations

Lecture 10

A Simple Example of the Construction of an Extremal Field; Rayleigh Quotients and the Method of Rayleigh-Ritz

We now consider, in order to have a simple example of the construction of an extremal field, the plane problem of Type 1 of the calculus of variations whose Lagrangian function is

$$L = \frac{1}{2}\left\{c\, x_t^2 + 2\, d\, x_t + e\, x^2 + 2\, f\, x\right\}$$

where the coefficients c, d, e, f, are either constants or, possibly, functions of t. The Lagrangian momentum is

$$p = c\, x_t + d$$

and the matrix $L_{x^*_t x_t}$, here 1 - dimensional, is c. The Euler-Lagrange equation is

$$c\, x_{tt} + c_t\, x_t + d_t = e\, x + f$$

and we assume, in order to be assured of the uniqueness as well as the existence of a solution of this equation for which the initial values of x and x_t are arbitrarily prescribed, that c and d

possess continuous second derivatives,and that e and f possess continuous first derivatives over a given open interval $t_0 < t < t_1$. The problem is positively regular over any region D of our 3 - dimensional state-space whose projection on the (t, x)-plane is covered by the strip $t_0 < t < t_1$ of this plane, which strip is parallel to the x - axis if c is positive over the interval $t_0 < t < t_1$, and, when this is the case, the strong Legendre condition is satisfied at every point of every extremal curve covered by the strip $t_0 < t < t_1$. Let $x = x(t)$, $a \leqslant t \leqslant b$, where $t_0 < a < b < t_1$, be an arc C of any extremal curve covered by the strip $t_0 < t < t_1$, and let $\bar{x} = \bar{x}(t)$, $a \leqslant t \leqslant b$ be any piecewise-smooth curve $\bar{C}$, which is covered by the strip $t_0 < t < t_1$ and which has the same end-points as C. Writing $\bar{x}(t) = x(t) + y(t)$, so that both $y(a)$ and $y(b)$ are zero, we have

$$I(\bar{C}) = \frac{1}{2} \int_a^b \left\{ c\,(x_t + y_t)^2 + 2\,d\,(x_t + y_t) + e\,(x + y)^2 + 2\,f\,(x + y) \right\} dt$$

and it follows, since

$$I(C) = \frac{1}{2} \int_a^b \left\{ c\,x_t^2 + 2\,d\,x_t + e\,x^2 + 2\,f\,x \right\} dt$$

that

$$I(\overline{C}) - I(C) = \int_a^b \left\{ (cx_t + d)\, y_t + \frac{1}{2} c y_t^2 + (ex + f)\, y + \frac{1}{2} e y^2 \right\} dt$$

Upon integration by parts we obtain, since both $y(b)$ and $y(a)$ are zero,
$\int_a^b (c\, x_t + d)\, y_t \, dt = - \int_a^b (c\, x_t + d)_t \, y \, dt$ and, since C is an extremal curve, this may be written as $- \int_a^b (ex + f)\, y \, dt$. Thus

$$I(\overline{C}) - I(C) = \frac{1}{2} \int_a^b (c\, y_t^2 + e\, y^2)\, dt$$

and, if $e \geqslant 0$ over the interval $a \leqslant t \leqslant b$, the right-hand side of this equation is $\geqslant \frac{1}{2} \int_a^b c\, y_t^2 \, dt$ which is, in turn, $\geqslant 0$.

The weak inequality $\geqslant$ may be replaced by the strong inequality $>$ unless $y_t = 0$ at all its points of continuity. Since $y(a) = 0$, this would imply that $y = 0$ over the interval $a \leqslant t \leqslant b$. Thus the inequality $e \geqslant 0$ over $t_0 < t < t_1$ is sufficient to ensure that $I(\overline{C}) > I(C)$, where C is any arc of any extremal curve which is covered by the strip $t_0 < t < t_1$, and $\overline{C}$ is any piecewise-smooth curve which is covered by this strip and has the same end-points as C (it being understood that $\overline{C}$ is furnished by an equation of the

form $\bar{x} = \bar{x}(t)$, $a \leqslant t \leqslant b$, so that no two points of $\bar{C}$ have the same projection on the t-axis). The strong inequality $I(\bar{C}) > I(C)$ assures us that there does not exist more than one extremal curve connecting any two given points $A : (a, x(a))$ and $B : (b, x(b))$ of the strip $t_0 < t < t_1$; indeed, if two such extremal curves C_1 and C_2 existed, we would have on taking C_1 as C and C_2 as $\bar{C}$ the inequality $I(C_2) > I(C_1)$, and we would also have on taking C_2 as C and C_1 as $\bar{C}$ the contradictory inequality $I(C_1) > I(C_2)$. Thus any arc of any extremal curve which is covered by the strip $t_0 < t < t_1$ is a minimal curve, and no other curve connecting the end-points of this arc is a minimal curve. Moreover, this arc of the extremal curve is not only a minimal curve but it yields an absolute minimum of the integral $I = \int_a^b L\, dt$.

The condition $e \geqslant 0$, $t_0 < t < t_1$, while sufficient to yield the strong inequality $I(\bar{C}) > I(C)$, where $\bar{C}$ is any piecewise-smooth curve of the form $\bar{x} = \bar{x}(t)$, $a \leqslant t \leqslant b$, with the same end-points as C, is not necessary for the validity of this inequality. For example, let us consider the case where the coefficients, c, d, e, and f are all constant, c being 1, d and f being zero, and e being $-k^2$ where $k > 0$. Our Euler-Lagrange equation is

$x_{tt} + k^2x = 0$, so that our extremals are of the form $x = s \cos(kt - \delta)$, where s and δ are constants of integration. We take as our given end-points the points $\begin{pmatrix} -1 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and as our region D of our 3 - dimensional state-space the slice $-1 - \epsilon < t < 1 + \epsilon$, perpendicular to the t - axis, where ϵ is a positive number which we may take to be as small as we please. On setting $\delta = 0$ we obtain the 1 - parameter family of extremal curves $x = s \cos kt$, all of which intersect the x - axis. In order that this 1 - parameter family of extremal curves be an extremal field over the strip $-1 - \epsilon < t < 1 + \epsilon$, all that is necessary (since $n = 1$) is that it simply cover this strip, and the condition for this is that the Jacobian matrix of $\begin{pmatrix} x \\ t \end{pmatrix}$ with respect to $\begin{pmatrix} s \\ t \end{pmatrix}$ be nonsingular over the strip. The determinant of this Jacobian matrix is $x_s = \cos kt$ and so the 1 - parameter family of extremal curves $x = s \cos kt$ will be an extremal field over the strip $-1 - \epsilon < t < 1 + \epsilon$ if, and only if, $k(1 + \epsilon) < \frac{\pi}{2}$ or, equivalently, since ϵ is arbitrarily small if, and only if, $k < \frac{\pi}{2}$. (If $k \geq \frac{\pi}{2}$ all the extremal curves of the family $x = s \cos kt$ have the common points $\begin{pmatrix} \pm \frac{\pi}{2k} \\ 0 \end{pmatrix}$, so that the family does not simply cover the strip $-1 - \epsilon < t < 1 + \epsilon$

no matter how small the positive number ϵ). When $k < \frac{\pi}{2}$,

so that the 1 - parameter family of extremals $x = s \cos kt$ is an extremal field over the strip $-1-\epsilon < t < 1+\epsilon$ if ϵ is sufficiently small, the field function $w(t, x)$ is $-k x \tan k t$, since $x_t = k s \sin k t = -k x \tan k t$. Since $L = \frac{1}{2}(x_t^2 - k^2 x^2)$,

we have $P_1 = L - px_t = -\frac{1}{2}(x_t^2 + k^2 x^2)$, $P_2 = p = x_t$ so that

$$P(X, W(X))\, X_\tau = -\frac{1}{2} k^2 x^2 (\tan^2 kt + 1) - k x \tan k t\, x_t$$

(our parameter τ being the Lagrangian parameter t).

Hence, the Weierstrass E-function is furnished by the formula

$$\begin{aligned} E(X, x_t, w(X)) = L - P(X, W(X))\, X_t &= \frac{1}{2}(x_t^2 + k^2 x^2 \tan^2 kt) + kx \tan kt\, x_t \\ &= \frac{1}{2}(x_t + kx \tan kt\, x_t)^2 \end{aligned}$$

so that

$$I(\overline{C}) - I(C) = \int_a^b E(\overline{X}, \bar{x}_t, w(\overline{X}))\, dt = \frac{1}{2}\int_a^b (\bar{x}_t + k\,\bar{x} \tan kt\, \bar{x}_t)^2\, dt \geqslant 0 .$$

The weak inequality $\geqslant$ may be replaced by the strong inequality $>$ when $\overline{C}$ does not coincide with C, for the equality would imply that $\bar{x}_t + k\,\bar{x} \tan kt\, \bar{x}_t = 0$ at all the smooth points of $\overline{C}$, and this would, in turn, imply that $\bar{x} = \bar{s} \cos kt$, where $\bar{s}$ is a constant, at all the smooth points of $\overline{C}$. Since $\cos kt \neq 0$ over the strip $-1-\epsilon < t < 1+\epsilon$,

if ϵ is sufficiently small, and since $x = 0$ at the end-points of our extremal, the constant of integration $\bar{s}$ must be zero, so that $\bar{C}$ must coincide with C. The member C of our extremal field $x = s \cos kt$, $-1-\epsilon < t < 1+\epsilon$, which passes through the two points $\begin{pmatrix}-1\\0\end{pmatrix}$ and $\begin{pmatrix}1\\0\end{pmatrix}$, is furnished by setting $s = 0$, this member being the line segment connecting these two points; thus x and x_t, and consequently L are zero along C, so that $I(C) = 0$. Hence, $I(\bar{C}) > 0$, it being always understood that $0 < k < \frac{\pi}{2}$. This striking result may be stated as follows:

If $\bar{x} = \bar{x}(t)$, $-1 \leqslant t \leqslant 1$, is any piecewise-smooth curve connecting the two points $\begin{pmatrix}-1\\0\end{pmatrix}$, $\begin{pmatrix}1\\0\end{pmatrix}$, other than the line segment which connects these points, the integral $\int_{-1}^{1} \bar{x}_t^2\, dt$ is greater than k^2 times the integral $\int_{-1}^{1} \bar{x}^2\, dt$, where k is any positive number $< \frac{\pi}{2}$; or, equivalently, if $\bar{x} = \bar{x}(t)$, $-1 \leqslant t \leqslant 1$, is any piecewise-smooth curve connecting the two points $\begin{pmatrix}-1\\0\end{pmatrix}$, $\begin{pmatrix}1\\0\end{pmatrix}$, other than the line segment which connects these two points, the number $\frac{\pi^2}{4}$ is $\leqslant$ the quotient
$$\frac{\int_{-1}^{1} \bar{x}_t^2\, dt}{\int_{-1}^{1} \bar{x}^2\, dt}.$$

Thus $\frac{\pi^2}{4}$ is a lower bound for the collection of such quotients. That it is the greatest lower bound of this collection becomes clear on taking $\bar{x} = A \cos \frac{\pi}{2} t$, $-1 \leqslant t \leqslant 1$, where A is any constant, since, for this smooth curve which connects the two points $\begin{pmatrix} -1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$, the quotient in question is $\frac{\pi^2}{4}$. We may regard, if we wish, this property of the number $\frac{\pi^2}{4}$ as a definition of the number π, and we may use this definition to appraise the number π. Thus if we set, for example,

$$\bar{x} = 1 - t^2, \quad -1 \leqslant t \leqslant 1, \quad \text{the quotient} \quad \frac{\int_{-1}^{1} \bar{x}_t^{\,2}\, dt}{\int_{-1}^{1} x^2\, dt} = 2.5,$$

and we know that $\frac{\pi^2}{4} \leqslant 2.5$.

Actually $\frac{\pi^2}{4} = 2.4674$ to 4 decimal places, so that the appraisal is astonishingly good.

Our Lagrangian function $L = \frac{1}{2} (x_t^{\,2} - k^2 x^2)$ may be regarded as the Lagrangian function of a mechanical system,

with one degree of freedom, for which the kinetic energy $T = \frac{1}{2} x_t^2$ and the potential energy $V = \frac{1}{2} k^2 x^2$.

The Euler-Lagrange equation of this problem, namely, $x_{tt} + k^2 x = 0$, has, when $k > 0$ the general solution $x = s \cos (k t - \delta)$, and for this curve to pass through the two points $\begin{pmatrix} -1 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ without being the line segment connecting these two points, we must have $\cos (k + \delta) = 0$, $\cos (k - \delta) = 0$. These two equations are equivalent to the two equations $\cos k \cos \delta = 0$, $\sin k \sin \delta = 0$. From the first of these two equations we see that either k or δ must be an odd integral multiple of $\frac{\pi}{2}$, and from the second we see that, if δ is an odd integral multiple of $\frac{\pi}{2}$, then k is a nonzero integral multiple of π. In either event, k is a nonzero integral multiple of $\frac{\pi}{2}$, and we term the squares of these nonzero integral multiples of $\frac{\pi}{2}$ the characteristic numbers of the boundary-value problem furnished by the differential equation $x_{tt} + \lambda x = 0$ and the boundary conditions $x(-1) = 0$, $x(1) = 0$. The least of these characteristic numbers is $\frac{\pi^2}{4}$, and we have seen that this least characteristic number is the

greatest lower bound of the collection of quotients

$$R = \frac{\int_{-1}^{1} x_t^2 \, dt}{\int_{-1}^{1} x^2 \, dt}$$

for all piecewise-smooth curves of the form $x = x(t)$, $-1 \leqslant t \leqslant 1$, which connect the two points $\begin{pmatrix} -1 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$, the line segment that connects these two points being excluded. We denote any one of this collection of quotients by the symbol R, since the connection between these quotients and the least characteristic number of certain boundary-value problems, whose characteristic numbers are all positive, was first stated by Rayleigh; and we term R the Rayleigh quotient for the piecewise-smooth function $x = x(t)$, $-1 \leqslant t \leqslant 1$, which is arbitrary except that it must satisfy the boundary conditions without being identically zero.

Let us now consider a family of piecewise-smooth functions $x = x(t, s)$, $-1 \leqslant t \leqslant 1$, depending on a parameter, or parametric matrix s, the only care necessary in the choice of this family being that each of its members must satisfy the boundary conditions Then the Rayleigh quotient $R = R(s)$ for any member of the

family which is not identically zero will be a function of the parameter, or parametric matrix s. No matter what the value of s, $R(s)$ will be $\geqslant$ the least characteristic number we are seeking ($\frac{\pi^2}{4}$ in the simple example we have chosen to illustrate the method). By selecting s so as to minimize $R(s)$, we obtain the best approximation to the least characteristic number we are seeking which we can obtain by the use of the family we have adopted. This modification of Rayleigh's procedure was proposed by Ritz, and the method is known as the Rayleigh-Ritz method. For example, in the problem we are considering, we may set $x(t) = (1 - t^2)(1 + s t^2)$, the factor $1 - t^2$ automatically taking care of the boundary conditions and the other factor being taken to be $1 + s t^2$ rather than $1 + s t$, in view of the symmetry of the problem about $t = 0$. A simple calculation yields $R(s) = \frac{3}{2}\left(\frac{35 + 14 s + 11 s^2}{21 + 6 s + s^2}\right)$, and this has a minimum value when s is the greater of the two roots of the quadratic equation $26 s^2 + 196 s + 42 = 0$. Taking $s = -0.22$, which is this root to two decimal places, we find $R(s) = 2.467438$ to 6 decimal places and this is greater than $\frac{\pi^2}{4} = 2.467401$ by less than 4 units in the fifth decimal place. Observe that the

minimizing value of the parameter s does not have to be determined with great precision, since the Rayleigh quotient $R(s)$, the quantity in which we are interested, is stationary at this minimizing value.

Lectures on Applied Mathematics

The Calculus of Variations

Lecture 11

The Principle of Maupertuis; The Propagation of Waves

We have seen how to pass from the Hamiltonian function $H(X, p)$ of a problem of Type 1 of the calculus of variations to the Lagrangian function $L(X, x_t)$ of this problem; all we have to do is to solve the equation $H_p(X, p) = x_t$ for p as a function of X and x_t and set $L = p\,x_t - H(X, p)$. In order to present this passage in parametric form, we observe that the role of the Hamiltonian function $H(X, p)$ is to furnish, through the equation

$$\phi(Z) \equiv P_1 + H(X, p) = 0$$

the $(2N - 1)$ - dimensional point locus, in the extended phase space whose points are $Z = \begin{pmatrix} X \\ P \end{pmatrix}$, on which lie the images of all piecewise-smooth curves in our N - dimensional time-coordinate space. The particular form in which the equation of this $(2N - 1)$ - dimensional point locus is written is of no significance, and it is sometimes convenient, for reasons of symmetry, to change this

form (without changing, naturally, the point locus itself). For example, for the arc-length problem in the plane,

$$L = (1 + x_t^2)^{1/2}, \quad p = \frac{x_t}{(1 + x_t^2)^{1/2}}, \quad x_t = \frac{p}{(1 - p^2)^{1/2}},$$

$H(X, p) = p\,x_t - L = -(1 - p^2)^{1/2}$, so that $\phi(Z) \equiv P_1 - (1 - P_2^2)^{1/2}$.

The point-locus in the 4 - dimensional extended phase space whose equation is $\phi(Z) \doteq P_1 - (1 - P_2^2)^{1/2} = 0$ may be more symmetrically described by the equation $(Z) = \frac{1}{2}(P_1^2 + P_2^2 - 1) = 0$.

We write any such equation of our $(2N - 1)$ - dimensional point locus as $\psi(Z) = 0$, where $\psi(Z)$ is taken to be a continuously differentiable function of Z. The gradient ψ_Z of ψ is a multiple, in general varying with Z, of the gradient ϕ_Z of ϕ, and this latter gradient is $(H_X; 1, H_p)$. The relation $p = L_{x_t}$, which defines the Lagrangian momentum matrix p, is equivalent to the relation $x_t = H_p$, and so the last N of the elements of ϕ_Z are those of $X_t = X_\tau \tau_t$.

Hence the last N of the elements of ψ_Z are proportional to those of X_τ, and we write

$$X_\tau = \lambda \psi_P$$

where λ is a multiplier which depends on the particular equation $\psi(Z) = 0$ we have chosen to represent our $(2N - 1)$ - dimensional point locus in our $2N$ - dimensional extended phase space, λ being t_τ when $\psi = \Phi$ We term the function $\psi(Z)$ a parametric Hamiltonian function, and in order to pass from any parametric Hamiltonian function ψ to the corresponding parametric integrand $F = F(X, X_\tau)$, we proceed as follows:

1) We solve the $N + 1$ equations

$$\lambda \psi_P = X_\tau \; ; \; \psi(Z) = 0$$

for P and λ as functions of X and X_τ. In order to be assured of the possibility of carrying out this step, we assume that the $(N + 1)$ - dimensional matrix, $\begin{pmatrix} \psi_{PP*} & \psi_P \\ \psi_{P*} & 0 \end{pmatrix}$ exists and is continuous and nonsingular over a region of our $2N$ - dimensional extended phase space which covers the closed point set in the space furnished by the equation (Z) = 0. The determinant of the $(N + 1)$ - dimensional Jacobian matrix of $\begin{pmatrix} \lambda \psi_P \\ \psi \end{pmatrix}$ with respect to $\begin{pmatrix} P* \\ \lambda \end{pmatrix}$, namely, $\begin{pmatrix} \lambda \psi_{PP*} & \psi_P \\ \psi_{P*} & 0 \end{pmatrix}$, is the product of the

determinant of the $(N+1)$ - dimensional matrix $\begin{pmatrix} \psi_{PP*} & \psi_P \\ \psi_{P*} & 0 \end{pmatrix}$

by λ^{n-1}, and so our hypothesis assures us of the possibility of solving, without ambiguity, the $N+1$ equations $\lambda\psi_P = X_\tau$, $\psi(Z) = 0$ for P and λ as functions of X and X_τ at all points where X_τ is not the zero $N \times 1$ matrix (so that $\lambda \neq 0$). The function, $\lambda = \lambda(X, X_\tau)$, which we obtain in this way is a positively homogeneous function, of degree 1, of X_τ, whereas the elements of the $1 \times N$ matrix P are positively homogeneous functions, of degree zero, of X_τ, since, from any solution (P, λ) of our $N+1$ equations, we obtain a solution $(P, k\lambda)$ of the $N+1$ equations $\lambda\psi_P = k X_\tau$, $\psi(Z) = 0$. Note: The adjective "positively" is necessary, since λ may be determined by means of an algebraic equation of degree higher than 1. For example, for the arc-length problem in the plane, if we take $\psi(Z)$ to be $\frac{1}{2}\left(P_1^2 + P_2^2 - 1\right)$, our equations are $\lambda P_1 = t_\tau$, $\lambda P_2 = x_\tau$, $P_1^2 + P_2^2 = 1$, and $\lambda^2 = t_\tau^2 + x_\tau^2$. Taking $\lambda = \left(t_\tau^2 + x_\tau^2\right)^{1/2}$, we obtain a function of X_τ which is

positively homogeneous of degree 1 but not homogeneous of degree 1.

2) We set $\mathbf{F}(\mathbf{X}, \mathbf{X}_\tau) = \mathbf{P}\,\mathbf{X}_\tau$, thus obtaining a positively homogeneous function, of degree 1, of $\mathbf{X}_\tau$. The function $\mathbf{F}(\mathbf{X}, \mathbf{X}_\tau)$, which we obtain in this way, is the parametric integrand of a calculus of variations problem of Type 1 of which $\psi(\mathbf{Z})$ is a parametric Hamiltonian function. Indeed $\mathbf{F}_{\mathbf{X}_\tau} = \mathbf{P} + \mathbf{X}^*_\tau\, \mathbf{P}^*_{\mathbf{X}_\tau} = \mathbf{P}$, since $\mathbf{P}$ is a positively homogeneous function, of degree zero, of $\mathbf{X}_\tau$. The integral $\mathbf{I} = \int_\alpha^\beta \mathbf{F}(\mathbf{X}, \mathbf{X}_\tau)\, d\tau$ may then be presented as the line integral $\int \mathbf{P}\, d\mathbf{X}$ in our 2N - dimensional extended phase space, and the equations of the extremals of the constrained problem of Type 1 of the calculus of variations which is furnished by this integral and the constraint $\psi(\mathbf{Z}) = 0$ are

$$\frac{d\,\mathbf{X}^j}{\psi_{\mathbf{P}_j}} = \frac{d\,\mathbf{P}_k}{-\psi_{\mathbf{X}^k}} = \lambda\, d\tau, \quad j, k = 1, \ldots, N.$$

All that remains, then, in order to justify our statement that $\psi(\mathbf{Z})$ is a parametric Hamiltonian function of the calculus of variations problem of Type 1, whose parametric integrand is $\mathbf{F}$, is to show that $\mathbf{P}_\tau = -\lambda\,\psi_{\mathbf{X}}$ along any extremal of this problem or,

equivalently, to show that $F_X \quad - \lambda \psi_X$. On differentiating the equation $\psi(Z) = 0$ with respect to X, X_T (not P) being held constant, we obtain $\psi_X + \psi_{P^*} P^*_X = 0$ or, equivalently, $\lambda\psi_X + X^*_T \quad P^*_X = 0$. Since $F = P X_T = X^*_T \; P^*$, $F_X = X^*_T P^*_X$, so that $F_X = -\lambda\psi_X$.

<u>Example</u>. We consider the case of a conservative mechanical system with n degrees of freedom for which $T - \frac{1}{2} x^*_t \, G \, x_t$, where the n dimensional symmetric matrix G is a function of x alone, not involving t, and $V = V(x)$ is also a function of x alone. Then $L = \frac{1}{2} x^*_t \, G \, x_t \quad V$, $p = x^*_t \, G$, $x_t = G^{-1} p^*$, it being understood that G is nonsingular. Thus $T = \frac{1}{2} p \, G^{-1} p^*$ and $H(x, p) = \frac{1}{2} p \, G^{-1} p^* + V$. Our $(2N-1)$ - dimensional point locus in the $2N$ - dimensional extended phase space has the equation $P_1 + \frac{1}{2} p \, G^{-1} p^* + V = 0$ and, since $H_t = 0$, we know that along any given extremal curve of our calculus of variations problem, H has a constant value, E say, where E, the energy of the conservative mechanical system, varies, naturally, from one extremal curve to another. In order to discuss an extremal curve for which the value of E is given, we may write $\frac{1}{2} p \, G^{-1} p^* + V - E = 0$, since $P_1 = -H$, and treat the left-hand side

of this equation as a parametric Hamiltonian function of a calculus of variations problem whose extended phase space is $2N-2$ dimensional, rather than $2N$ - dimensional, any point of the new extended phase space being furnished by a $2n \times 1$ matrix $\begin{pmatrix} x \\ p \end{pmatrix}$, rather than a $2N \times 1$ matrix $\begin{pmatrix} X \\ P \end{pmatrix}$. When we proceed in this way we say that we have ignored the time coordinate, and we may similarly ignore any or all coordinates which do not appear in the Hamiltonian function $H(X, p)$, the corresponding constant momentum, or momenta, playing the role of the energy constant E.

We now ask, what is the parametric integrand of the calculus of variation problem of Type 1 of which $\frac{1}{2} p G^{-1} p^* + V - E$ is a parametric Hamiltonian function? Denoting this function by $\psi\begin{pmatrix} x \\ p \end{pmatrix}$, we write down the $n+1$ equations $\lambda \psi_p = x_\tau, \psi\begin{pmatrix} x \\ p \end{pmatrix} = 0$, i.e.,

$$\lambda G^{-1} p^* = x_\tau ; \quad \frac{1}{2} p G^{-1} p^* + V - E = 0$$

The first n of these yield $p^* = \frac{1}{\lambda} G x_\tau$ and we obtain, on substituting this in the last equation, $\lambda^2 = \frac{1}{2(E-V)} x^*_\tau G x_\tau$.

Hence

$$F = p x_\tau = \frac{x^*_\tau G x_\tau}{\lambda} = \left\{2(E-V)\right\}^{1/2} \left\{x^*_\tau G x_\tau\right\}^{1/2}$$

Thus those extremal curves of our original calculus of variations problem for which the value E of the energy $H = T + V$ is given are the extremal curves of a new calculus of variations problem whose parametric integrand is the function F just written. If we assign to the coordinate space of our mechanical system the metric which is defined by the formula $(ds)^2 = dx^* G\, dx$, then F and $I = \int_{\alpha}^{\beta} F\, d\tau$ appear in the form

$$F = \left\{2(E-V)\right\}^{1/2} s_{\tau} \; ; \quad I = \int_{S_1}^{S_2} \left\{2(E-V)\right\}^{1/2} ds$$

and our new calculus of variations problem is closely related to the arc-length problem in n - dimensional space. If the mechanical system were free from applied force, so that V is constant, our new calculus of variations problem is precisely the arc-length problem in n - dimensional space, and the paths of the mechanical system would be the geodesics in the n - dimensional coordinate space, always on the understanding that this space is assigned the metric which is defined by the formula $(ds)^2 = (dx)^* G\, dx$. When V is not constant, the paths of the mechanical system are extremal curves related to the integral

$$I = \int n\, ds \; ; \qquad n = 2\left\{(E-V)\right\}^{1/2}$$

Integrals such as this occur in the theory of the propagation of light through a refracting isotropic medium, n being known as the index of refraction. Thus the problem of determining those paths of a conservative mechanical system for which the energy has an assigned constant value E is the same as that of determining the rays along which light is propagated in a medium whose index of refraction is proportional to $(E-V)^{1/2}$, the metric assigned to the medium being furnished by the formula $ds^2 = dx^* G\, dx$, where G is the matrix of the homogeneous quadratic form in x_t which furnishes the kinetic energy of the conservative mechanical system. This connection between mechanics and optics was pointed out more than two centuries ago by Maupertuis, and is known as the Principle of Maupertuis.

We now consider briefly the propagation of light in a refracting isotropic medium whose index of refraction, $n = n(x)$, is a function of only one of the three space coordinates; the adjective isotropic means that $n = n(x)$ does not depend on x_s, and when, in addition, n involves only one of the three space coordinates, we refer to the medium as stratified. If we are using rectangular Cartesian coordinates (x, y, z), and if

$n = n(z)$ does not involve the coordinates x and y, we say that the medium is stratified in planes perpendicular to the z - axis, while, if we are using space polar coordinates (r, θ, ϕ), and $n = n(r)$ does not involve the coordinates θ and ϕ, we say that the medium is stratified in concentric spheres centered at the origin. This latter case is of importance in the discussion of the propagation of earthquake waves through the interior of the earth (the index of refraction n being a constant multiple of the reciprocal of the velocity of propagation of the waves). Whether the problem is one dealing with the propagation of light or of elastic waves, the integral $I = \int n\, ds$ is a constant multiple of the time taken for the disturbance to pass from one point of the medium to another, so that the theory is governed, in each case, by Fermat's Principle of Least Time.

Treating first the case of a plane stratified medium, the parametric momentum matrix is furnished by the formulas $P_1 = n x_s$; $P_2 = n y_s$; $P_3 = n z_s$; and $P_1^2 + P_2^2 + P_3^2 = n^2$. The coordinates x and y are ignorable, P_1 and P_2 being constants along any extremal. Setting $P_1 = c_1$, $P_2 = c_2$, we have $n^2 z_s^2 = (n^2 - c_1^2 - c_2^2)$, so that, along any extremal, s is furnished, in terms of z, by the quadrature,

$$s = \int^{z} \frac{n\,dz}{(n^2 - c_1^2 - c_2^2)^{1/2}}\,.$$

It is clear, since $y_x = \dfrac{P_2}{P_1} = \dfrac{c_2}{c_1}$ if $c_1 \neq 0$, and $x_y = \dfrac{c_1}{c_2}$ if $c_2 \neq 0$, that, if not both c_1 and c_2 are zero, the corresponding extremal lies in a plane parallel to the z - axis, and there is no loss of generality in taking this plane to be the plane $y = 0$, so that $c_2 = 0$; when we do this we write c instead of c_1, so that $s = \displaystyle\int^{z} \frac{n\,dz}{(n^2 - c^2)^{1/2}}$ along our extremal. If both c_1 and c_2 are zero, the extremal is a straight line segment parallel to the z - axis. Note: The reason that we can obtain the extremals without integrating the one nontrivial Euler-Lagrange equation $(n\,z_s)_s = n_z$ is that our calculus of variations problem is presented in parametric form; the function $\psi(Z) = \frac{1}{2}(P_1^2 + P_2^2 + P_3^2 - n^2)$ is a parametric Hamiltonian function and the combination of the two trivial Euler-Lagrange equations $(n\,x_s)_s = 0$, $(n\,y_s)_s = 0$, and the relation $\psi(Z) = 0$ may be used to replace the three Euler-Lagrange equations once the constant values of $P_1 = n\,x_s$ and $P_2 = n\,y_s$ along any extremal are given.

It follows from the relation $n^2 z_s^2 = n^2 - c_1^2 - c_2^2$ that $n^2 (1 - z_s^2) = c_1^2 + c_2^2$, so that $n (1 - z_s^2)^{1/2}$ is constant along any extremal; this result, which states that the sine of the angle, at any point of the extremal, between the extremal and the z-axis is inversely proportional to the index of refraction at that point, is known as Snell's Law of Refraction. If the index of refraction is discontinuous at $z = z_1$ (a finite number of such discontinuities being allowed by the theory) the extremals have corners at $z = z_1$, but $P_3 = n (1 - z_s^2)^{1/2}$ is continuous at $z = z_1$.

The velocity of a mass particle falling under gravity is proportional to $z^{1/2}$, where z is the depth below the level of zero velocity, and so the brachistochrone problem, i.e., the problem of determining the curve connecting two points along which the time of descent is a minimum, is the same as that of determining the curves along which light is propagated in an isotropic medium stratified in planes perpendicular to the z - axis, the index of refraction being proportional to $z^{-1/2}$. The velocity v of the mass particle is furnished by the formula $v^2 = 2 g z$, where g is the acceleration due to

gravity, and so we set $v = (2gz)^{1/2}$. Assuming that our extremal is not a line segment parallel to the z - axis and that it lies in the (z, x) - plane, we have

$$s = \int_0^z \frac{dz}{(1-2gc^2z)^{1/2}} = \frac{1}{gc^2}\left\{1-(1-2gc^2z)^{1/2}\right\};\quad c \neq 0,\ s \leq \frac{1}{gc^2}$$

s being measured from the level of zero velocity $z = 0$. The maximum value of z is $\frac{1}{2gc^2}$, and when z reaches this value it begins to decrease, so that s_z is negative and is furnished by the formula $s_z = -(1-2gc^2z)^{-1/2}$, rather than by the formula $s_z = (1-2gc^2z)^{-\frac{1}{2}}$. We denote $\frac{1}{2gc^2}$ by $2a$, so that the length of the descending part of the curve, from $z = 0$ to $z = 2a$, is $\frac{1}{gc^2} = 4a$. On denoting by ϕ the angle made by the curve with the z - axis, we have $\cos\phi = z_s = \pm(1-2gc^2z)^{1/2}$, ϕ varying from 0 to $\frac{\pi}{2}$ over the descending part and from $\frac{\pi}{2}$ to π over the ascending part of the curve. Thus $s = 4a(1-\cos\phi)$ over both the descending and the ascending part of the curve, and $z_\phi = z_s s_\phi = 4a \sin\phi\cos\phi$, so that $z = a(1-\cos 2\phi)$,

the constant of integration being determined by the fact that $z = 0$ when s, or, equivalently ϕ, is 0. Finally, $x_s = \sin\phi$, so that $x_\phi = x_s s_\phi = 4a\sin^2\phi$ and $x = x_0 + a(2\phi - \sin 2\phi)$, where x_0 is the value of x when $\phi = 0$. Thus the parametric equations of any extremal curve other than a line segment parallel to the z-axis, are

$$z = a(1-\cos 2\phi)\,;\; x = x_0 + a(2\phi - \sin 2\phi)\,;\; 0 \leqslant \phi \leqslant \pi$$

These extremal curves are cycloids traced by a point on the circumference of a circle of radius $\underline{a}$ rolling on and below the level of zero velocity. This family of extremal curves is a 2-parameter family, the two parameters being a and x_0, and it can be shown, by a detailed discussion into which we do not enter, that there passes through any two points whose z-coordinates are nonnegative and whose x-coordinates are different, one, and only one, member of the family. Furthermore, if we fix the value of x_0, the one-parameter family so obtained simply covers the region $z > 0$, $x > x_0$ of the (z, x)-plane and furnishes an extremal field over this region. The problem of Type 1 of the calculus of variations which is furnished by the integral $I = \int n\,ds$, $n > 0$ is, since the arc-length problem is positively

regular, itself positively regular, and so we know that if our two given points lie below the level of zero velocity and do not have the same x - coordinate, the one and only cycloid of our 2 - parameter family which connects them furnishes an absolute minimum of the integral I. The excluded cases, namely, those when one or both of the two given points lie on the level of zero velocity and when both have the same x - coordinate, are easily cared for (the first by an argument involving considerations of continuity and the second by using as our extremal field the 1 - parameter family of straight lines parallel to the z - axis).

For a spherically stratified medium the parametric momentum matrix is $P_1 = n\, r_s$, $P_2 = n\, r^2\, \theta_s$, $P_3 = n\, r^2 \sin^2\theta\, \phi_s$, the parameter adopted being the arc-length, and the second and third of these are constant along any extremal $n\, r^2\, \theta_s = c_2$, $n\, r^2 \sin^2\theta\, \phi_s = c_3$. There is no loss of generality in taking $c_3 = 0$, as this can be arranged by properly choosing the space polar coordinate reference frame. Thus the extremal curves are plane curves, and we may confine our attention to plane polar coordinates (r, θ) writing, simply, $n\, r^2\, \theta_s = c$.

$$\frac{1}{2}\left(P_1^2 + \frac{1}{r^2} P_2^2 - n^2\right)$$

is a parametric Hamiltonian function, and, since $P_2 = c$, it

follows that $n\,r\,(1 - r_s^2)^{1/2}$ is constant along any extremal.

Thus the product by r of the sine of the angle which the extremal makes, at any of its points, with the radius vector is inversely proportional to the index of refraction at that point, this being the analogue, for a spherically stratified medium, of Snell's Law of Refraction for a plane stratified medium. The relation $P_1^2 + \frac{c^2}{r^2} = n^2$ yields $n^2 r_s^2 + \frac{c^2}{r^2} = n^2$, so that, along any extremal,

$$s = \int_{r_0}^{r} \frac{n\,r\,dr}{(n^2 r^2 - c^2)^{1/2}},$$ s being measured from $r = r_0$. For example, if n is inversely proportional to r, s is a constant times $r - r_0$ along any extremal. For any given value of c, $n\,r \geqslant c$ along the corresponding extremal; for example, if $n = (2gr)^{-1/2}$, the nearest point of the extremal to the center of the spherically stratified medium has a distance $= 2\,gc^2$ from this center.

Lectures on Applied Mathematics

The Calculus of Variations

Lecture 12

Problems Whose Lagrangian Functions Involve Derivatives of Higher Order than the First

We shall consider in this lecture problems of Type 1 of the calculus of variations whose Lagrangian functions involve derivatives of higher order than the first. Confining our attention, at the beginning, to the case where L involves second-order derivatives, but no derivatives of order higher than the second, L is a function of the following three matrices

1) The $N \times 1$ time - coordinate matrix $X = \begin{pmatrix} t \\ x \end{pmatrix}$.

2) The $n \times 1$ velocity matrix x_t.

3) The $n \times 1$ acceleration matrix x_{tt}.

We denote by z the $(3n + 1) \times 1$ matrix $\begin{pmatrix} X \\ x_t \\ x_{tt} \end{pmatrix}$ rather than, as before, the $(2n + 1) \times 1$ matrix $\begin{pmatrix} X \\ x_t \end{pmatrix}$. Our state space is now $(3n + 1)$ - dimensional rather than $(2n + 1)$ - dimensional, and the $(2n + 1)$ - dimensional space whose points are furnished

by the $(2n+1) \times 1$ matrices $\begin{vmatrix} X \\ x_t \end{vmatrix}$, i.e., our previous state space, will now play the role previously played by the N - dimensional time-coordinate space By this we mean that the various curves $x = x(t)$, $a \leqslant t \leqslant b$ in our time-coordinate space will not only have their end-points $A = \begin{pmatrix} a \\ x(a) \end{pmatrix}$ and $B = \begin{pmatrix} b \\ x(b) \end{pmatrix}$ prescribed, but also the end values $x_t(a)$ and $x_t(b)$ of the velocity matrix will be prescribed. It no longer suffices that these curves be piecewise smooth, since the acceleration matrix x_{tt} must be defined along them, except, possibly, at a finite number of points. We say that a curve $x = x(t)$, $a \leqslant t \leqslant b$ in our time-coordinate space has piecewise-continuous curvature if it possesses the following two properties:

1) It is smooth, i.e., x_t exists and is continuous over $a \leqslant t \leqslant b$

2) The interval $a \leqslant t \leqslant b$ may be covered by a finite net over each cell of which x_{tt} exists and is continuous. If t' is an interior point of the net the right-hand and left-hand acceleration matrices at t', $x_{tt}(t'+0)$ and $x_{tt}(t'-0)$, exist, but they need not be equal. If they are equal x_{tt} exists and is continuous at t', and the point t' may be removed from the net.

Each point of a curve in our time-coordinate space possessing

piecewise-continuous curvature at which x_{tt} exists, defines a

point $z = \begin{pmatrix} X \\ x_t \\ x_{tt} \end{pmatrix}$ of our new $(3n+1)$ - dimensional state space,

and we confine our attention to those curves whose images in this state space are covered by a region D of this space over which $L = L(z)$ is, by hypothesis, a continuously differentiable function of the $(3n+1) \times 1$ matrix z If we wish to present our problem parametrically, we must assume that the function $t = t(\mathcal{T})$, $\alpha \leqslant \mathcal{T} \leqslant \beta$, which introduces the new parameter $\mathcal{T}$, is smooth over $\alpha \leqslant \mathcal{T} \leqslant \beta$ and, in addition, that this interval may be covered by a finite net, over each cell of which $t_{\mathcal{T}\mathcal{T}}$ exists and is continuous. Then

the curve $X = X(\mathcal{T})$, $\alpha \leqslant \mathcal{T} \leqslant \beta$, possesses piecewise-continuous curvature, and the parametric integrand $F(X, X_{\mathcal{T}}, X_{\mathcal{T}\mathcal{T}})$ is defined by the formula.

$$F(X, X_{\mathcal{T}}, X_{\mathcal{T}\mathcal{T}}) = L(X, x_t, x_{tt})\, t_{\mathcal{T}} : \; x_t = \frac{x_{\mathcal{T}}}{t_{\mathcal{T}}} \; ; x_{tt} = \frac{t_{\mathcal{T}} x_{\mathcal{T}\mathcal{T}} - x_{\mathcal{T}} t_{\mathcal{T}\mathcal{T}}}{t_{\mathcal{T}}^3}$$

In order to obtain the Euler-Lagrange equation we imbed any curve C possessing piecewise-continuous curvature and whose image in our $(3n+1)$ - dimensional state space is covered by D, in a 1 parameter family of curves $\bar{C}$, all possessing piecewise-continuous curvature and all of whose images are covered by D,

as follows If C is furnished by the formula

$$\mathbf{X} = \mathbf{X}(\tau), \quad \alpha \leqslant \tau \leqslant \beta$$

we write

$$\bar{\mathbf{X}} = \mathbf{X}(\tau) + s\, f(\tau)\ ; \quad \alpha \leqslant \tau \leqslant \beta\ ; \quad -\delta \leqslant s \leqslant \delta$$

where the $N \times 1$ matrix function $f(\tau)$ possesses the following two properties:

1) $f(\tau)$ possesses, over $\alpha \leqslant \tau \leqslant \beta$, a piecewise-continuous second derivative.

2) Both $f(\tau)$ and its first derivative, $f_\tau(\tau)$, vanish at $\tau = \alpha$ and at $\tau = \beta$.

These properties ensure that all our comparison curves $\bar{C}$ possess piecewise-continuous curvature and, moreover, that

$$\bar{\mathbf{X}}(\alpha) = \mathbf{X}(\alpha), \quad \bar{\mathbf{X}}_\tau(\alpha) = \mathbf{X}_\tau(\alpha), \quad \bar{\mathbf{X}}(\beta) = \mathbf{X}(\beta), \quad \bar{\mathbf{X}}_\tau(\beta) = \mathbf{X}_\tau(\beta),$$

so that all four $N \times 1$ matrices $\bar{\mathbf{X}}(\alpha)$, $\bar{\mathbf{X}}_\tau(\alpha)$, $\bar{\mathbf{X}}(\beta)$, $\bar{\mathbf{X}}_\tau(\beta)$ are independent of s. Since $\bar{\mathbf{X}}_\tau = \mathbf{X}_\tau + s\, f_\tau$, $\bar{\mathbf{X}}_{\tau\tau} = \mathbf{X}_{\tau\tau} + s\, f_{\tau\tau}$, we have $\delta\, \mathbf{X}_\tau = (\delta\, \mathbf{X})_\tau$, as in our previous discussion, and, also, $\delta\, \mathbf{X}_{\tau\tau} = (\delta\, \mathbf{X}_\tau)_\tau$. If δ is taken sufficiently small, all the curves $\bar{C}$ of our 1 - parameter family of comparison curves are such that their images in our $(3n + 1)$ - dimensional

state-space are covered by the region $\mathbf{D}$, over which $\mathbf{L} = \mathbf{L}(\mathbf{z})$ is, by hypothesis, a continuously differentiable function of $\mathbf{z}$ or, equivalently, over which $\mathbf{F}(\mathbf{X}, \mathbf{X}_\tau, \mathbf{X}_{\tau\tau})$ is a continuously differentiable function of the $3\,N \times 1$ matrix $\begin{pmatrix} \mathbf{X} \\ \mathbf{X}_\tau \\ \mathbf{X}_{\tau\tau} \end{pmatrix}$. The integral,

$$\mathbf{I} = \int_\alpha^\beta \mathbf{F}(\overline{\mathbf{X}}, \overline{\mathbf{X}}_\tau, \overline{\mathbf{X}}_{\tau\tau})\, d\tau,$$ of $\mathbf{F}$ along any one of the curves $\overline{\mathbf{C}}$ is a differentiable function of s and

$$\delta\,\mathbf{I} = \int_\alpha^\beta (\mathbf{F}_\mathbf{X}\,\delta\,\mathbf{X} + \mathbf{F}_{\mathbf{X}_\tau}\,\delta\,\mathbf{X}_\tau + \mathbf{F}_{\mathbf{X}_{\tau\tau}}\,\delta\,\mathbf{X}_{\tau\tau})\, d\tau$$

As in our previous discussion we may integrate the matrix product $\mathbf{F}_\mathbf{X}\,\delta\,\mathbf{X}$ by parts, obtaining

$$\int_\alpha^\beta (\mathbf{F}_\mathbf{X}\,\delta\,\mathbf{X})\, d\tau = -\int_\alpha^\beta \left\{ G\,(\delta\mathbf{X})_\tau \right\} d\tau, \text{ where } G(\tau) = \int_\alpha^\tau \mathbf{F}_\mathbf{X}(u)\, d\,u\,,$$

$\delta\,\mathbf{X}$ being zero at $\tau = \alpha$ and at $\tau = \beta$. Since $(\delta\mathbf{X})_\tau = \delta\mathbf{X}_\tau$, we have

$$\delta\,\mathbf{I} = \int_\alpha^\beta \left\{ \left(\mathbf{F}_{\mathbf{X}_\tau} - G\right) \delta\,\mathbf{X}_\tau + \mathbf{F}_{\mathbf{X}_{\tau\tau}}\,\delta\,\mathbf{X}_{\tau\tau} \right\} d\tau$$

and we now integrate by parts the matrix product, $\left(\mathbf{F}_{\mathbf{X}_\tau} - G\right) \delta\,\mathbf{X}_\tau$.

Setting $H(\tau) = \int_\alpha^\tau \left(\mathbf{F}_{\mathbf{X}_\tau}(u) - G\,(u)\right) d\,u$, we have

$$\int_\alpha^\beta \left\{(F_{X_\tau} - G)\, \delta X_\tau\right\} d\tau \;=\; -\int_\alpha^\beta \left\{H\,(\delta X_\tau)_\tau\right\} d\tau$$

since δX_τ is zero at $\tau = \alpha$ and at $\tau = \beta$. It follows, since

$(\delta X_\tau)_\tau = \delta X_{\tau\tau}$, that

$$\delta I = \int_\alpha^\beta \left[\left\{F_{X_{\tau\tau}} - H(\tau)\right\} \delta X_{\tau\tau}\right] d\tau$$

and it is easy to see that δI will be zero for all allowable

choices of $\delta X_{\tau\tau} = d\, s\, f_{\tau\tau}$, if $F_{X_{\tau\tau}} - H(\tau)$ is a linear

$1 \times N$ matrix function of τ. Indeed, on writing for our later convenience, any linear $1 \times N$ matrix function of τ in the form $(\tau - \alpha)\, c + d$, where c and d are constant $1 \times N$ matrices,

we have $\left\{(\tau - \alpha)\, c + d\right\} \delta X_{\tau\tau} = \sum_{j=1}^{N} \left\{(\tau - \alpha) c_j + d_j\right\} \delta X^j_{\tau\tau}$ and

1) $$\int_\alpha^\beta (\tau - \alpha) c_j\, \delta X^j_{\tau\tau}\, d\tau = c_j\, (\tau - \alpha)\, \delta X^j_\tau \Big|_\alpha^\beta - c_j \int_\alpha^\beta \delta X^j_\tau\, d\tau$$

$$= c_j\, (\tau - \alpha)\, \delta X^j_\tau \Big|_\alpha^\beta - c_j\, \delta X^j \Big|_\alpha^\beta$$

which is zero, since both δX^j and δX^j_τ are zero at $\tau = \alpha$ and at $\tau = \beta$; and

2) $\int_{\alpha}^{\beta} d_j\, \delta X^j_{\tau\tau} d\tau = d_j\, \delta X^j_{\tau} \Big|_{\alpha}^{\beta} = 0$

Note: In deriving these results we have made use of the relations

$\delta X^j_{\tau\tau} = (\delta X^j_{\tau})_{\tau}$, $\delta X^j_{\tau} = (\delta X^j)_{\tau}$.

Conversely, if δI is zero for all allowable choices of $\delta X_{\tau\tau}$, the $1 \times N$ matrix $F_{X_{\tau\tau}} - H(\tau)$ must be a linear function of τ. To see this, we introduce the $1 \times N$ matrix function of τ:

$$K(\tau) = \int_{\alpha}^{\tau} \left[\int_{\alpha}^{u} \left\{ F_{X_{\tau\tau}}(v) - H(v) \right\} dv \right] du - \frac{1}{2}(\tau - \alpha)^2 c - \frac{1}{6}(\tau - \alpha)^3 d$$

where c and d are constant $1 \times N$ matrices which we shall shortly determine. $K(\tau)$ is continuously differentiable over the interval $\alpha \leqslant \tau \leqslant \beta$, $K_{\tau}(\tau)$ being furnished by the formula:

$$K_{\tau}(\tau) = \int_{\alpha}^{\tau} \left\{ F_{X_{\tau\tau}}(v) - H(v) \right\} dv - (\tau - \alpha)c - \frac{1}{2}(\tau - \alpha)^2 d$$

Thus both of the $1 \times N$ matrices $K(\tau)$ and $K_{\tau}(\tau)$ are zero at $\tau = \alpha$, no matter what are the constant $1 \times N$ matrices c and d, and we now determine these constant $1 \times N$ matrices so that both of the $1 \times N$ matrices $K(\tau)$ and $K_{\tau}(\tau)$ are also zero at $\tau = \beta$.

This will be the case if, and only if, c_j and d_j, $j = 1, \ldots, N$, have the values which are unambiguously determined by the two linear equations

$$\frac{1}{2}(\beta - \alpha)^2 c_j + \frac{1}{6}(\beta - \alpha)^3 d_j = \int_\alpha^\beta \left[\int_\alpha^u \left\{ F_{X^j_{\tau\tau}}(v) - H_j(v) \right\} dv \right] du$$

$$(\beta - \alpha) c_j + \frac{1}{2}(\beta - \alpha)^2 d_j = \int_\alpha^\beta \left\{ F_{X^j_{\tau\tau}}(v) - H_j(v) \right\} dv \, .$$

$K(\tau)$ has, over the interval $\alpha \leqslant \tau \leqslant \beta$, a piecewise-continuous second derivative, $K_{\tau\tau}(\tau) = F_{X_{\tau\tau}}(\tau) - H(\tau) - c - (\tau - \alpha) d$, and so we may take for our $N \times 1$ matrix $f(\tau)$ the transpose $K^*(\tau)$ of the $1 \times N$ matrix $K(\tau)$: when we do this, δI, which may be written in the form

$$\delta I = \int_\alpha^\beta \left[\left\{ F_{X_{\tau\tau}} - H(\tau) - c - (\tau - \alpha) d \right\} \delta X_{\tau\tau} \right] d\tau$$

appears as the product of the integral, over the interval $\alpha \leqslant \tau \leqslant \beta$, of the squared magnitude of the $1 \times N$ matrix, $F_{X_{\tau\tau}} - H(\tau) - c - (\tau - \alpha) d$, by ds. Hence, for δI to be zero, this $1 \times N$ matrix must be zero at all its points of continuity; in other words, $F_{X_{\tau\tau}} - H(\tau) = c + (\tau - \alpha) d$ at all the points of

the curve $X = X(\tau)$, $\alpha \leqslant \tau \leqslant \beta$, at which $X_{\tau\tau}$ is defined (and, hence, by hypothesis continuous), c and d being unambiguously determinate constant 1 x N matrices. Since $H(\tau)$ is continuous over $\alpha \leqslant \tau \leqslant \beta$, it follows that if τ', say, is any one of the finite number of points of this interval at which $X_{\tau\tau}$ is, possibly, undefined, $F_{X_{\tau\tau}}(\tau' - 0) = F_{X_{\tau\tau}}(\tau' + 0)$. Since

$F = L\, t_\tau$ and $x_{tt} = \dfrac{t_\tau x_{\tau\tau} - t_{\tau\tau} x_\tau}{t^3_\tau}$, the first element of $F_{X_{\tau\tau}}$ is the negative of the quotient of the matrix product, $L_{x_{tt}} x_t$ by t_τ, while the remaining n elements of $F_{X_{\tau\tau}}$ are those of

$\frac{1}{t_\tau} L_{x_{tt}}$ Hence since t_τ is, by hypothesis, continuous over the interval $\alpha \leqslant \tau \leqslant \beta$, we have

$$L_{x_{tt}}(t' - 0) = L_{x_{tt}}(t' + 0)\ ;\ t' = t(\tau')$$

If the n - dimensional matrix $L_{x^*_{tt} x_{tt}}$ exists and is continuous and nonsingular over the region D of our (3 n + 1) - dimensional state-space, x_{tt} is unambiguously determined along the curve $x = x(t)$, $a \leqslant t \leqslant b$ by $L_{x_{tt}}$, and so $x_{tt}(t' - 0) = x_{tt}(t' + 0)$;

BLANK PAGE

in other words, x_{tt} is not merely piecewise continuous, but actually continuous over the interval $a \leqslant t \leqslant b$.

At any point of the curve $X = X(\tau)$, $\alpha \leqslant \tau \leqslant \beta$, at which $X_{\tau\tau}$ is defined, $H(\tau)$ is differentiable with the derivative $H_{\tau}(\tau) = F_{X_{\tau}} - G(\tau)$, and so we obtain from the relation

$F_{X_{\tau\tau}} - H(\tau) = c + (\tau - \alpha)\, d$, the existence, over $\alpha \leqslant \tau \leqslant \beta$, of $\left(F_{X_{\tau\tau}}\right)_{\tau}$, with the value $F_{X_{\tau}} - G(\tau) + d$. Moreover, at all such points, $G(\tau)$ is differentiable with the derivative $G_{\tau}(\tau) = F_{X}$. If the n - dimensional matrix $L_{x^{*}_{tt} x_{tt}}$ exists and is not only continuous and nonsingular over D but also continuously differentiable over D, x_{ttt} exists and is continuous over the interval $a \leqslant t \leqslant b$, and so $\left(F_{X_{\tau}}\right)_{\tau}$ exists and is continuous over $\alpha \leqslant \tau \leqslant \beta$ (it being necessary to assume, at this point, that $t_{\tau\tau\tau}$ exists and is continuous over $\alpha \leqslant \tau \leqslant \beta$). Hence $\left(F_{X_{\tau\tau}}\right)_{\tau\tau}$ exists over $\alpha \leqslant \tau \leqslant \beta$ with the value $\left(F_{X_{\tau}}\right)_{\tau} - F_{X}$, so that

$$\left(F_{X_{\tau\tau}}\right)_{\tau\tau} - \left(F_{X_{\tau}}\right)_{\tau} + F_{X} = 0\ ; \quad \alpha \leqslant \tau \leqslant \beta$$

This is the Euler-Lagrange equation for a problem of Type 1 of the calculus of variations whose Lagrangian function involves x_{tt}.

When we use the Lagrangian parameter t, this Euler-Lagrange equation appears as

$$\left(L_{x_{tt}}\right)_{tt} - \left(L_{x_t}\right)_t + L_x = 0 \qquad a \leqslant t \leqslant b$$

It is clear from the argument by which it was derived that if L involved, in addition to the matrices X, x_t, and x_{tt} the matrix x_{ttt}, the corresponding Euler-Lagrange equation would be

$$\left(L_{x_{ttt}}\right)_{ttt} - \left(L_{x_{tt}}\right)_{tt} + \left(L_{x_t}\right)_t - L_x = 0; \qquad a \leqslant t \leqslant b$$

and so on.

<u>Example</u>. $n = 1$, $L = \frac{1}{2} d\, x_{tt}^2 + \frac{1}{2}(e_2 x_t^2 + 2\, e_1 x_t) + \frac{1}{2}(f_2 x^2 + 2\, f_1 x)$, the coefficients d, e_2, e_1, f_2, and f_1 being given functions of t. $L_{x_{tt}} = d\, x_{tt}$ and the matrix $L_{x^*_{tt} x_{tt}}$, here 1 - dimensional, is the coefficient d. We assume, then, that d is continuously differentiable over $a < t < b$, and we know that our problem does not possess extremals unless $d\, x_{tt}$ possesses a continuous second derivative with respect to t along them. We assume, therefore, that not only is d continuously differentiable over $a \leqslant t \leqslant b$ but that it possesses a continuous second derivative over this interval.

Similarly, since $L_{x_t} = e_2\, x_t + e_1$, $L_x = f_2\, x + f_1$, we assume that e_2 and e_1 are continuously differentiable, and that f_2 and f_1 are continuous over $a \leqslant t \leqslant b$. Our Euler-Lagrange equation:

$$(d\, x_{tt})_{tt} - (e_2\, x_t + e_1)_t + f_2 + f_1 = 0$$

is an ordinary linear differential equation of the fourth order. Writing $\bar{x}\,(t) = x\,(t) + y\,(t)$, $a \leqslant t \leqslant b$, where both $y\,(t)$ and $y_t(t)$ are zero at $t = a$ and at $t = b$, we find

$$I\,(\bar{C}) - I\,(C) = \int_a^b \Big\{ d\,(x_{tt}\, y_{tt} + \tfrac{1}{2}\, y_{tt}^{\;2}) + e_2\,(x_t\, y_t + \tfrac{1}{2}\, y_t^{\;2})$$
$$+ e_1\, y_t + f_2\,(xy + \tfrac{1}{2}\, y^2) + f_1\, y \Big\}\, d\,t$$

A repeated integration by parts yields, since both $y\,(t)$ and $y_t\,(t)$ are zero at $t = a$ and at $t = b$,

$$\int_a^b d\, x_{tt}\, y_{tt}\, d\,t = -\int_a^b (d\, x_{tt})_t\, y_t\, d\,t = \int_a^b (d\, x_{tt})_{tt}\, y\, dt$$

and, since $\int_a^b (e_2\, x_t + e_1)\, y_t\, d\,t = -\int_a^b (e_2\, x_t + e_1)_t\, y\, dt$, we have

$$I\,(\bar{C}) - I\,(C) = \int_a^b \Big\{ (d\, x_{tt})_{tt} - (e_2\, x_t + e_1)_t + f_2\, x + f_1 \Big\}\, y\, dt$$
$$+ \int_a^b \tfrac{1}{2} \Big\{ d\, y_{tt}^2 + e\, y_t^2 + y^2 \Big\}\, d\,t$$

Since C is, by hypothesis, an extremal curve of our problem, the first of the two integrals on the right is zero, since its integrand is zero, and so

$$I(\overline{C}) - I(C) = \frac{1}{2} \int_a^b (dy_{tt}^2 + e_2 y_t^2 + f_2 y^2)\, dt$$

Thus, if d, e_2, and f are nonnegative over $a \leqslant t \leqslant b$, $I(\overline{C}) \geqslant I(C)$.

If d is positive over $a \leqslant t \leqslant b$, the weak inequality $I(\overline{C}) \geqslant I(C)$ may be replaced by the strong inequality $I(\overline{C}) > I(C)$ unless y_{tt} is zero at all its points of continuity; if this were the case y would be zero over $a \leqslant t \leqslant b$, since both y and y_t are zero at $t = a$. Hence, the relations $d > 0$, $e_1 \geqslant 0$, $f_2 \geqslant 0$ over $a \leqslant t \leqslant b$ are sufficient to ensure that $I(\overline{C}) > I(C)$, and consequently, that our extremal curve is a minimal curve, and moreover, that it is the one and only extremal curve for which $x(t)$ and $x_t(t)$ have prescribed values at $t = a$ and at $t = b$.

When presented parametrically, our integrand function $F(X, X_\tau, X_{\tau\tau})$ is such that $F(X, kX_\tau, k^2 X_{\tau\tau}) = k\, F(X, X_\tau, X_{\tau\tau})$, $k > 0$, and we obtain on differentiating this relation with respect to k and then setting $k = 1$:

$$P X_\tau + 2 Q X_{\tau\tau} = F \quad ; \quad P = F_{X_\tau}, \quad Q = F_{X_{\tau\tau}}$$

On denoting $L_{x_{tt}}$ by q, we have $Q_1 \; - q\dfrac{x_\tau}{t_\tau^{\,2}}$, $Q_2 = \dfrac{q_1}{t_\tau}$,

$Q_N = \dfrac{q_n}{t_\tau}$, so that $Q X_\tau = 0$. When the end-points of our curves may vary, we have

$$\delta I = \left\{ (P - Q_\tau)\,\delta X + Q\,\delta X_\tau \right\}\Big|_A^B + \int_\alpha^\beta \left\{ (F_X - F_{X_\tau} + F_{X_{\tau\tau}})\,\delta X \right\} d\tau$$

so that, when the curve of integration is an extremal curve,

$$\delta I = \left\{ (P - Q_\tau)\,\delta X + Q\,\delta X_\tau \right\}\Big|_A^B$$

The Hilbert invariant integral is

$$I^* = \int \left\{ (P - Q_\tau)\, d X + Q\, d X_\tau \right\}$$

where the two $N \times 1$ matrix field-functions V, U, are substituted for $X_{\tau\tau}$ and $X_{\tau\tau\tau}$, respectively, in the coefficients $P - Q_\tau$ and Q. When the curve of integration is an arc of a member of our extremal field (now a $2n$ - parameter field) I^* has the same value as I. Indeed, since $Q X_\tau = 0$, $F = P X_\tau + 2 Q X_{\tau\tau}$ may be written as $(P - Q_\tau) X_\tau + Q X_{\tau\tau}$, so that

$$I = \int_\alpha^\beta F d\tau = \int_\alpha^\beta \left\{ (P - Q_\tau) X_\tau + Q X_{\tau\tau} \right\} d\tau$$

As an example, let us consider the problem for which $L = (x_{tt}^2 - k^4 x^2)$, k a positive constant. The Euler-Lagrange equation is $x_{tttt} - k^4 x = 0$, so that the extremal curves are of the form $x = s_1 \cos(kt - \delta_1) + s_2 \cosh(kt - \delta_2)$, where s_1, s_2, δ_1, δ_2 are constants of integration. The 2 - parameter family obtained by setting $\delta_1 = 0$, $\delta_2 = 0$ is an extremal field over the slice $-1 - \epsilon < t < 1 + \epsilon$ of our $\begin{pmatrix} X \\ x_t \end{pmatrix}$ space if the Jacobian matrix of

$$\begin{pmatrix} s_1 \cos kt + s_2 \cosh kt \\ -ks_1 \sin kt + ks_2 \sinh kt \end{pmatrix}$$ with respect to the 2×1 matrix

$$\begin{pmatrix} s_1 \\ s_2 \end{pmatrix}$$ is nonsingular over this slice, i.e., if

$\cos kt \sinh kt + \sin kt \cosh kt$ is positive over the interval $-1 - \epsilon < t < 1 + \epsilon$. The least positive zero of the equation $\cos k \sinh k + \sin k \cosh k = 0$ plays the role played by $\frac{\pi}{2}$, the least positive zero of the equation $\cos k = 0$, in the problem whose Lagrangian function was $\frac{1}{2}(x_t^2 - k^2 x^2)$. As in that problem, we form the Rayleigh quotients

$$R = \frac{\int_{-1}^{1} x_{tt}^2 \, dt}{\int_{-1}^{1} x^2 \, dt}$$

for functions $x = x(t)$, $-1 \leqslant t \leqslant 1$, which vanish, together with their first derivatives, at $t = -1$ and at $t = 1$. Then the fourth power of the least positive zero of the equation $\cos k \sinh k + \sin k \cosh k = 0$ is the greatest lower bound of all the Rayleigh quotients which we obtain in this way. In order to justify this statement we have to show that

$\int_{-1}^{1} (\bar{x}_{tt}^2 - k^4 \bar{x}^2)\, dt$ is positive for all curves, other than the line segment $\bar{x}(t) = 0$, which are such that $\bar{x}(t)$ and $\bar{x}_t(t)$ are zero at $t = -1$ and at $t = 1$, provided that k is any positive number less than the least positive zero of the equation

$\cos k \sinh k + \sin k \cosh k = 0$. In order to show that $I(\overline{C}) > I(C) = 0$, we calculate the Weierstrass E - function,

$F(\overline{X}, \overline{X}_\tau, \overline{X}_{\tau\tau}) - (P - Q_\tau)\overline{X}_\tau - Q\overline{X}_{\tau\tau}$. Since

$$F = L\, t_\tau = \frac{1}{2}(x_{tt}^2 - k^4 x^2)\, t_\tau \quad \text{and} \quad x_{tt} = \frac{t_\tau x_{\tau\tau} - x_\tau t_{\tau\tau}}{c_\tau^3}, \text{we find}$$

$$P_1 = F_{t_\tau} = \frac{(t_\tau x_{\tau\tau} - x_\tau t_{\tau\tau})\, x_{\tau\tau}}{t_\tau^5} - \frac{5\,(t_\tau x_{\tau\tau} - x_\tau t_{\tau\tau})^2}{2\, t_\tau^6} - \frac{1}{2} k^4 x^2$$

$$P_2 = F_{x_\tau} = - \frac{(t_\tau x_{\tau\tau} - x_\tau t_{\tau\tau})\, t_{\tau\tau}}{t_\tau^5};$$

$$Q_1 = F_{t_{\tau\tau}} = - \frac{(t_\tau x_{\tau\tau} - x_\tau t_{\tau\tau})\, x_\tau}{t_\tau^5}; \quad Q_2 = F_{x_{\tau\tau}} = \frac{t_\tau x_{\tau\tau} - x_\tau t_{\tau\tau}}{t_\tau^4}$$

Using the Lagrangian parameter t, $t_\tau = 1$, $t_{\tau\tau} = 0$. so that

$$P_1 = -\frac{3}{2} x_{tt}^2 - \frac{1}{2} k^4 x^2 \; ; \; P_2 = 0 \; ; \; Q_1 = - x_t x_{tt} \; ; \; Q_2 = x_{tt}$$

and the Weierstrass E - function appears as

$$E = \frac{1}{2} (\bar{x}_{tt}^2 - k^4 \bar{x}^2) + \frac{1}{2} (x_{tt}^2 + k^4 x^2) - x_t x_{ttt} + x_{ttt} \bar{x}_t - x_{tt} \bar{x}_{tt}$$

where $x_{tt} = - k^2 s_1 \cos kt + k^2 s_2 \cosh kt$; $x_{ttt} = k^3 s_1 \sin kt + k^3 s_2 \sinh kt$;

and s_1 and s_2 are determined unambiguously as functions of x and x_t by means of the equations $x = s_1 \cos kt + s_2 \cosh kt$, $x_t = - k s_1 \sin kt + k s_2 \sinh kt$. Since $x = \bar{x}$, $x_t = \bar{x}_t$, this reduces to $\frac{1}{2} \bar{x}_{tt}^2 + \frac{1}{2} x_{tt}^2 - x_{tt} \bar{x}_{tt} = \frac{1}{2} (\bar{x}_{tt} - x_{tt})^2$, and so $I(\bar{C}) - I(C) = \int_{-1}^{1} E \, dt \geqslant 0$, the equality holding only when $\bar{x}_{tt} = x_{tt}$ at all the points of $\bar{C}$ at which $\bar{x}_{tt}$ exists (and is, consequently, continuous). Since both $\bar{x} - x$ and $(\bar{x} - x)_t$ are zero at $t = \pm 1$, it follows that $I(\bar{C}) - I(C) = 0$ only when $\bar{C}$ coincides with C. Hence $I(\bar{C}) > I(C) = 0$ when $\bar{C}$ is different from C, and it follows that $\int_{-1}^{1} \bar{x}_{tt}^2 \, dt \Big/ \int_{-1}^{1} \bar{x}^2 \, dt > k^4$, where k is any positive number less than the smallest positive zero k^* of $\cos k \sinh k + \sin k \cosh k$, and $\bar{x}(t)$ is any function of t which

possesses a piecewise-continuous second derivative over the interval $-1 < t \leqslant 1$, and which vanishes, together with its first derivative, at $t = -1$ and at $t = 1$. If we set $\bar{x} = \cosh k^* \cos k^*t - \cos k^* \cosh k^*t$, we have $\bar{x}_{tt} = -k^{*2} (\cosh k^* \cos k^*t + \cos k^* \cosh k^*t)$, and, since

$$\int_{-1}^{1} \cos k^*t \cosh k^*t \, dt = \frac{1}{k^*} (\cos k^* \sinh k^* + \sin k^* \cosh k^*) = 0,$$

the Rayleigh quotient for this function is k^{*4}. Hence k^{*4} is the greatest lower bound of all the Rayleigh quotients

$\int_{-1}^{1} \bar{x}_{tt}^2 \, dt \Big/ \int_{-1}^{1} x^2 \, dt$. The Rayleigh quotient for the function, $\bar{x} = (1 - t^2)^2$, is $\frac{63}{2}$ and so the least positive zero of $\cos k \sinh k + \sin k \cosh k$ is $< (31.5)^{1/4} = 2.36907$. The equation $\cos k \sinh k + \sin k \cosh k = 0$ may be written in the equivalent form $\cos 2k \cosh 2k - 1 = 0$ and the least positive zero of $\cos 2k \cosh 2k - 1$ is $k^* = 2.36502$ to five decimal places (as given by Rayleigh, "Theory of Sound," Vol. 1, p. 223).

<u>Exercise</u>. Set $\bar{x} = (1 - t^2)^2 (1 + s t^2)$ and determine the minimum value of the Rayleigh quotient, $R(s) = \int_{-1}^{1} \bar{x}_{tt}^2 \, dt \Big/ \int_{-1}^{1} \bar{x}^2 \, dt$, thus obtaining a closer approximation to the fourth power of the least positive zero of $\cos 2k \cosh 2k - 1$.

Lectures on Applied Mathematics

The Calculus of Variations

Lecture 13

Multiple-Integral Problems of the Calculus of Variations

In problems of Type 1 of the calculus of variations, we were concerned with integrals $I = \int_a^b L\,dt$ extended along piecewise-smooth curves $x = x(t)$, $a \leqslant t \leqslant b$ in our N - dimensional time-coordinate space, x being a $n \times 1$ matrix and N being $n + 1$. We now turn to problems of the calculus of variations in which we are concerned with integrals $I = \int L\,d(x)$ extended over n-dimensional point loci in N $(= n + 1)$ - dimensional space, and we refer to these problems as problems of Type 2. Instead of n functions $x = x(t)$ of a single independent variable or parameter, we now have one function $u = u(x)$ of n independent variables or parameters, the elements of a $n \times 1$ matrix x. As the point x varies over some region, plus its boundary, of the n - dimensional parametric x-space, the equation $u = u(x)$ furnishes an n-dimensional point locus in the N - dimensional $\begin{pmatrix} x \\ u \end{pmatrix}$ - space. If $n = 2$, this is a surface in 3-dimensional space, and we shall denote it by the symbol S even when n is greater than 2. On denoting by $R + R'$ the region, plus

its boundary, of our n - dimensional x - space over which x varies, the equation of S is

$$u = u(x)\ ; \quad x \subset R + R'$$

We shall suppose that S is smooth, i.e., that u possesses, over $R + R'$, a continuous gradient $1 \times n$ matrix u_x. This gradient matrix plays the role played by the $n \times 1$ velocity matrix x_t in problems of Type 1. We denote by X the $N \times 1$ matrix $\begin{pmatrix} x \\ u \end{pmatrix}$ and by z the $(2n + 1) \times 1$ matrix $\begin{pmatrix} X \\ u_x \end{pmatrix}$, and we term the $(2n + 1)$ - dimensional z-space the state-space of our problem of Type 2 of the calculus of variations. The Lagrangian function $L = L(z) - L(X, u_x)$ of this problem is a function of $z = \begin{pmatrix} X \\ u_x \end{pmatrix}$, which we assume to be a continuous function of x and a continuously differentiable function of u and u_x over a given region D of our $(2n + 1)$ - dimensional state-space. If X is any point of a smooth n - dimensional point locus S in our N - dimensional X - space, the $(2n + 1) \times 1$ matrix $\begin{pmatrix} X \\ u_x \end{pmatrix}$ furnishes a point in our $(2n + 1)$ - dimensional state space, and we term the point locus consisting of all these points the image Γ, in the state-space, of S. If this image Γ is covered by D, the integral

$$I = \int_{R + R'} L(X, u_x)\, d(x) \quad ; \quad d(x) = d(x^1, \ldots, x^n)$$

exists by virtue of the continuity of $L(z)$ over D and of the smoothness of S. The points of S which are furnished by the boundary R' of R constitute the boundary of S, and we denote this $(n - 1)$ - dimensional boundary of S by S', and our problem of Type 2 of the calculus of variations has two parts:

1) The determination of the extremal n - dimensional point loci; i.e., the point loci for which the integral I, evaluated for all smooth n - dimensional point loci having a prescribed $(n - 1)$ - dimensional boundary, has a stationary value.

2) The determination of necessary, or of sufficient, conditions for an extremal n - dimensional point locus S to be a minimal n - dimensional point locus; i.e., such that $I(\bar{S}) \geqslant I(S)$ if $\bar{S}$ is any comparison n - dimensional point locus which is sufficiently near S.

Note: We shall, from now on, refer to an n-dimensional point locus in our N - dimensional $\begin{pmatrix} x \\ u \end{pmatrix}$ - space as a surface, even when $n > 2$. When $n = 1$, the distinction between problems of Type 1 and problems of Type 2 of the calculus of variations disappears.

In order to obtain the Euler-Lagrange equation for our problem of Type 2 of the calculus of variations, we imagine any smooth surface S whose equation is

$$u = u(x) \quad ; \quad x \subset R + R'$$

and whose image Γ in our $(2n+1)$- dimensional state-space is covered by D, as imbedded in a 1 - parameter family of smooth comparison surfaces $\bar{S}$ whose equations are

$$\bar{u} = u(x) + s\,f(x) \quad ; \quad x \subset R + R' \quad ; \quad -\delta \leqslant s \leqslant \delta$$

where $f(x)$ is continuously differentiable over $R + R'$ and is zero over R', so that all the members of our family of comparison surfaces have the same boundary as S. If δ is sufficiently small, the image $\bar{\Gamma}$ of any member $\bar{S}$ of this family of comparison surfaces in our $(2n+1)$ - dimensional state space is covered by D, and the integral

$$I(\bar{S}) = \int_{R+R'} L(\bar{X}, \bar{u}_x)\, d(x)$$

exists and is a differentiable function of s over the interval $-\delta \leqslant s \leqslant \delta$. We denote by δI the differential of this function of s, evaluated at $s = 0$, and we term δI the variation of I. Then

$$\delta I = \int_{R+R'} \left\{ L_u\, \delta u + \delta u_x\, L_{u_x} \right\} d(x)$$

where $\delta u = ds\, f$ and $\delta u_x = ds\, f_x$ are the variations of u and u_x, respectively, ds being an arbitrary number. L_{u_x}, which is an $n \times 1$ matrix, plays, for problems of Type 2 of the calculus of variations, the role played by the $1 \times n$ Lagrangian momentum matrix

$p = L_{x_t}$ for problems of Type 1. We observe that $\delta u_x = (\delta u)_x$ and we avail ourselves of this fact to transform, under the hypothesis that L_{u_x} is continuously differentiable over $R + R'$, the multiple integral $\int_{R+R'} \left\{ \delta u_x \, L_{u_x} \right\} d\,(x)$, just as, when treating problems of Type 1, we used the rule of integration by parts to transform the corresponding integral $\int_a^b \left\{ p\delta\,(x_t) \right\} dt$. The rule which corresponds for multiple integrals to the rule of integration by parts for simple integrals is as follows: If v and w are functions of the $n \times 1$ matrix x which are continuously differentiable over $R + R'$, then

$$\int_{R+R'} v\, w_{x^1}\, d\,(x) = \int_{R'} v\, w\, d\,(x^2, \ldots, x^n) - \int_{R+R'} v_{x^1}\, w\, d\,(x)$$

$$\int_{R+R'} v\, w_{x^2}\, d\,(x) = -\int_{R'} v\, w\, d\,(x^1, x^3, \ldots, x^n) - \int_{R+R'} v_{x^2}\, w\, d\,(x)$$

$$\cdots\cdots\cdots\cdots\cdots$$

$$\int_{R+R'} v\, w_{x^n}\, d\,(x) = (-1)^{n-1} \int_{R'} v\, w\, d\,(x^1, \ldots, x^{n-1}) - \int_{R+R'} v_{x^n}\, w\, d\,(x)$$

If $w = \begin{pmatrix} w^1 \\ \vdots \\ w^n \end{pmatrix}$ is a $n \times 1$ matrix function of the $n \times 1$ matrix x which is continuously differentiable over $R + R'$, it follows that

$$\int_{R+R'} v \operatorname{div} w \, d(x) - \int_{S'} v w \, d S' - \int_{R+R'} v_x w \, d(x)$$

where div w, the divergence of w, is the sum,

$w^1_{x^1} + \ldots + w^n_{x^n}$, of the diagonal elements of the n - dimensional

matrix w_x, and $w \, d S'$ is an abbreviation for

$$d(x^2, \ldots, x^n) w^1 - d(x^1, x^3, \ldots, x^n) w^2 + \ldots + (-1)^{n-1} d(x^1, \ldots, x^{n-1}) w^n$$

evaluated over R'. The integral $\int_{S'} v w \, d S'$ is known as the integral,

over the boundary S' of S of the n x 1 matrix v w, or as the

flux, through S', of the vector which is furnished by the n x 1

matrix v w. If v is zero over R', we obtain the simplified rule

$$\int_{R+R'} v \operatorname{div} w \, d(x) = -\int_{R+R'} v_x w \, d(x)$$

Setting $v = \delta u$, $w = L_{u_x}$, under the hypothesis that L_{u_x} is

continuously differentiable over R + R', we obtain

$$\int_{R+R'} (\delta u)_x L_{u_x} \, d(x) = -\int_{R+R'} (\delta u \operatorname{div} L_{u_x}) \, d(x)$$

and it follows, since $(\delta u)_x = \delta u_x$, that

$$\delta I = \int_{R+R'} \left\{ \left(L_u - \operatorname{div} L_{u_x} \right) \delta u \right\} d(x)$$

The same argument as in the case of problems of Type 1 shows that the necessary and sufficient condition that δI be zero, for all allowable choices of $\zeta(x)$, is that $u = u(x)$ satisfy, over $R + R'$, the partial differential equation

$$L_u - \operatorname{div} L_{u_x} = 0$$

which is the Euler-Lagrange equation for our problem of Type 2 of the calculus of variations. Indeed, if $L_u - \operatorname{div} L_{u_x}$ were different from zero at a single point of $R + R'$ it would, by virtue of its continuity, be different from zero at a point x^* say, of R and hence it would be, again by virtue of its continuity, one-signed over a neighborhood of x^* which is covered by R. Taking $\zeta(x)$ to be one-signed over this neighborhood and zero elsewhere, δI could not be zero, since it reduces to the integral of the product by ds of the function $(L_u - \operatorname{div} L_{u_x}) \zeta(x)$ over the neighborhood in question, and this function is one-signed over this neighborhood. This proves the necessity of the equation $L_u - \operatorname{div} L_{u_x} = 0$, $x \subset R + R'$, for δI to be zero for all allowable choices of $\zeta(x)$, and the sufficiency of this condition is evident.

Example. Denoting by $\Delta_1 u$ the squared magnitude, $u_x u_x^*$, of the gradient $1 \times n$ matrix u_x of u:

$$\Delta_1 u = u_x u_x^* = (u_{x^1})^2 + \ldots + (u_{x^n})^2$$

let $L = \Delta_1 u + g u^2$, where g is a given function of x which is continuous over $R + R'$. Then $L_{u_x} = 2 u_x^*$, so that $\operatorname{div} L_{u_x} = 2 \Delta_2 u$, where $\Delta_2 u = u_{x^1 x^1} + \ldots + u_{x^n x^n}$ is the sum of the diagonal elements of the n - dimensional matrix $u_{x^* x}$, i.e. the Laplacian of u. Since $L_u = 2 g u$, the Euler-Lagrange equation is

$$\Delta_2 u = g u$$

and the problem of determining extremal surfaces of the Lagrangian function $L = \Delta_1 u + g u^2$ over a given region R of our n - dimensional x-space is that of determining a solution of the second-order partial differential equation $\Delta_2 u = g u$, which takes assigned values over the boundary R' of R.

Lectures on Applied Mathematics

The Calculus of Variations

Lecture 14

Constrained Problems; Characteristic Numbers

Since for problems of Type 2 of the calculus of variations we have only one dependent variable, we cannot, as we did when considering problems of Type 1, impose a constraint of the form $\phi\,(X) \equiv \phi\begin{pmatrix} x \\ u \end{pmatrix} = 0$, since a constraint of this type would determine u as a function of x and would not permit us to vary u. We may, however, impose a weaker constraint of the form

$$\int_{R+R'} \phi\,(X)\,d\,(x) = C\ ;\ \phi_u \text{ continuous and not identically zero}$$

over $R + R'$, where C is some given constant. On differentiating this relation with respect to s, when u is replaced by $\bar{u} = u + s\mathfrak{f}(x)$, and then setting $s = 0$, we obtain

$$\int_{R+R'} (\phi_u\ \delta\,u)\ d\,(x) = 0$$

and in order that $\delta\,I$, where $I = \int_R L\,(X,\ u_x)\ d\,(x)$, may be zero for all allowable choices of $\mathfrak{f}(x)$, it is necessary and sufficient that

$$\int_{R + R'} \left\{ \left(L_u - \operatorname{div} L_{u_x} \right) \delta u \right\} d(x) = 0$$ for all allowable choices of

$\delta u = f(x)\, ds$ which respect the relation $\int_{R + R'} (\phi_u \, \delta u) \, d(x) = 0$.

It is evident, then, that a sufficient condition for δI to be zero, for all allowable choices of $f(x)$, is that $L_u - \operatorname{div} L_{u_x}$ be a constant multiple of ϕ_u and the following argument shows that this sufficient condition is also necessary. If λ is any constant, the statement that $\int_{R + R'} \left\{ \left(L_u - \operatorname{div} L_{u_x} \right) \delta u \right\} d(x)$ is zero for all allowable choices of δu which respect the relation $\int_{R + R'} (\phi_u \, \delta u) \, d(x) = 0$ implies that $\int_{R + R} \left\{ \left(L_u - \operatorname{div} L_{u_x} - \lambda \phi_u \right) \delta u \right\} d(x) = 0$ for all allowable choices of δu which respect the relation $\int_{R + R'} (\phi_u \delta u) \, d(x) = 0$.

If we were allowed to set $\delta u = L_u - \operatorname{div} L_{u_x} - \lambda \phi_u$, the equation

$$\int_{R + R'} \left\{ \left(L_u - \operatorname{div} L_{u_x} - \lambda \phi_u \right) \delta u \right\} d(x) = 0$$ or, equivalently,

$$\int_{R + R'} \left(L_u - \operatorname{div} L_{u_x} - \lambda \phi_u \right)^2 d(x) = 0$$ would imply, since

$L_u - \operatorname{div} L_{u_x} - \lambda \phi_u$ is, by hypothesis, continuous over $R + R'$, that

$L_u - \operatorname{div} L_{u_x} = \lambda \phi_u$ over $R + R'$. In order to respect the condition $\int_{R+R'} \left\{ \phi_u \, \delta u \right\} d(x) = 0$, λ must be, when we set $\delta u = L_u - \operatorname{div} L_{u_x} - \lambda \phi_u$ the quotient of the integral of $(L_u - \operatorname{div} L_{u_x}) \phi_u$ over $R + R'$ by the integral of ϕ_u^2 over $R + R'$. There remains the difficulty that we are not allowed to set $\delta u = L_u - \operatorname{div} L_{u_x} - \lambda \phi_u$ without the assurance that $L_u - \operatorname{div} L_{u_x} - \lambda \phi_u$ is zero over R', but we avoid this difficulty as follows: Let R^*_ϵ be a subdomain of R which is such that $R^*_\epsilon \to R + R'$ as $\epsilon \to 0$. Then the integral of 1 over $R + R' - R^*_\epsilon$ tends to zero with ϵ, and we set $\delta u = L_u - \operatorname{div} L_{u_x} - \lambda \phi_u$, with the previously determined value of λ, over R^*_ϵ; whereas, over $R + R' - R^*_\epsilon$, δu is any continuous function of x which is zero over R'. Then the integral of $(L_u - \operatorname{div} L_{u_x} - \lambda \phi_u) \, \delta u$ over R^*_ϵ is nonnegative, whereas the integral of this same expression over $R + R' - R^*_\epsilon$ tends to zero with ϵ since the integrand is bounded

over $R + R' - R^*_\epsilon$. Since the integral of $(L_u - \text{div}\, L_{u_x} - \lambda \phi_u)\, \delta u$ over $R + R'$ is, by hypothesis, zero, the limit, as $\epsilon \to 0$, of the integral of this expression over R^*_ϵ is ze. , and since this integral does not decrease as $\epsilon \to 0$, its integrand being nonnegative, the value of the integral over R^*_ϵ must be ero ; this implies that $L_u - \text{div}\, L_{u_x} = \lambda \phi_u$ over R^*_ϵ ; regardless of the value of ϵ . It follows by virtue of the continuity over $R + R'$ of each of the two sides of this equation that $L_u - \text{div}\, L_{u_x} = \lambda \phi_u$ over $R + R'$.

The case where $\phi = \frac{1}{2} u^2$, so that $\phi_u = u$, is particularly interesting. If L does not involve u explicitly, so that $L_u = 0$, the partial differential equation which determines, in combination with the assigned values of u over R', the extremal surfaces of the constrained problem of Type 2 of the calculus of variations:

$$I = \int_{R + R'} L(x, u_x)\, d(x) \quad ; \quad \int_{R + R'} u^2\, d(x) = \text{constant}$$

is

$$\text{div}\, L_{u_x} + \lambda u = 0$$

and we see on multiplying this equation by u and integrating the product over $R + R'$ that λ is the negative of the quotient of the integral of $u \operatorname{div} L_{u_x}$ over $R + R'$ by the given value of $\int_{R+R'} u^2 \, d(x)$. Since $u = 0$ over R', the negative of the integral of $u \operatorname{div} L_{u_x}$ over $R + R'$ is the same as the integral over $R + R'$ of the matrix product $u_x L_{u_x}$. Taking the given value of $\int_{R+R'} \phi \, d(x)$ to be 1, so that $\int_{R+R'} u^2 \, d(x) = 2$, we see that

$$\lambda = \frac{1}{2} \int_{R+R'} (u_x L_{u_x}) \, d(x)$$

If, in particular, $L = \frac{1}{2} u_x u_{x*}$, so that $L_{u_x} = u_{x*}$, we have $u_x L_{u_x} = 2L$ (this relation being, indeed, valid when L is any homogeneous, or positively homogeneous, function of degree 2 of the $1 \times n$ matrix, u_x). Hence

$$\lambda = \int_{R+R'} L \, d(x) = I.$$

We shall now consider in some detail the particular constrained problem of Type 2 of the calculus of variations in which u is

assigned the value zero over R', and $L = \frac{1}{2} u_x u_x^*$. Then $\operatorname{div} L_{u_x} = \operatorname{Tr} u_{x*x} = \Delta_2 u$, and the extremal surfaces are determined by the partial differential equation:

$$\Delta_2 u + \lambda u = 0 \; ; \quad x \subset R + R'$$

and the boundary condition:

$$u = 0 \; ; \quad x \subset R'$$

The value of the constant λ, which is that of our i.tegral I extended over the extremal surface, is not given, and it is part of our problem to determine it. An obvious, but trivial, extremal surface is that furnished by the equation $u = 0$, $x \subset R + R'$, and for this extremal surface $\lambda = 0$. If there exists a nontrivial extremal surface, i.e., one for which u is not identically zero over $R + R'$, the corresponding value of λ is positive (since u is not constant over $R + R'$, its value over R' being zero, and so the integral of $L = \frac{1}{2} u_x u_x^*$ over our extremal surface is positive). We term any value of λ, for which the boundary-value problem

$$\Delta_2 u + \lambda u = 0, \quad x \subset R + R' ; \quad u = 0, \quad x \subset R'$$

possesses a nontrivial solution, a characteristic number of this boundary-value problem, so that all the characteristic numbers of this

particular boundary-value problem, if any such exist, are positive.

If our constrained problem of Type 2 of the calculus of variations possesses a minimal surface S, the corresponding boundary-value problem possesses a characteristic number, since the constraint $\frac{1}{2}\int_{R+R'} u^2\, d(x) = 1$ prohibits u from being identically zero over $R + R'$. Let us suppose that the minimal surface S furnishes an absolute minimum of the integral I for all functions u which are zero over R' and which respect the condition $\frac{1}{2}\int_{R+R'} u^2\, d(x) = 1$. If v is any function of x which is continuous over $R + R'$ and zero over R', we can determine the constant multiplier c so that $u = cv$ respects the condition $\frac{1}{2}\int_{R+R'} u^2\, d(x) = 1$, and so $\frac{1}{2}\int_{R+R'} (u_x^* u_x)\, d(x) \geqslant \lambda_1$ or, equivalently, $\frac{1}{2}\int_{R+R'} (v_x^* v_x)\, d(x) \geqslant \frac{1}{c^2}\lambda_1$

This implies, since $\frac{1}{2}\int_{R+R'} v^2\, d(x) = \frac{1}{c^2}$, that $\int_{R+R'} (v_x^* v_x - \lambda_1 v^2)\, d(x) \geqslant 0$. In other words, λ_1 is a lower

bound of the various Rayleigh quotients

$$R = \frac{\int_{R+R'} (v_x^* v_x)\, d(x)}{\int_{R+R'} v^2\, d(x)}$$

Example. Let R be the circle of radius a with center at the origin, and set $\lambda_1 = k_1^2$. When written in polar coordinates the equation $\Delta_2 u + \lambda_1 u = 0$ appears, on the assumption that u is a function of r alone, as $u_{rr} + \frac{1}{r} u_r + k_1^2 u = 0$, and the only solutions of this equation which are finite at $r = 0$ are constant multiples of $J_0(k_1 r)$, where J_0 is the Bessel function of index zero of the first kind. Thus $k_1 a$ is the smallest positive zero, 2.4048, to 4 decimal places, of this Bessel function, so that $\lambda_1 a^2 = 5.7832$

The Rayleigh quotient for $v = \cos \frac{\pi r}{2a}$ is the product of $\frac{\pi^2}{4a^2}$ by

$$\int_0^a (\sin^2 \frac{\pi r}{2a})\, rdr \div \int_0^a (\cos^2 \frac{\pi r}{2a})\, rdr = \int_0^a (1-\cos \frac{\pi r}{a})\, rdr \div \int_0^a (1+\cos\frac{\pi r}{a})\, rdr$$

$= \frac{\pi^2 + 4}{\pi^2 - 4}$, and hence the product of this Rayleigh quotient by a^2 is

$\frac{\pi^4 + 4\pi^2}{4(\pi^2 - 4)}$ = 5.8304, to 4 decimal places; this product being greater than the square of the least positive zero of J_0 by less than 5 units in the second decimal place.

Lectures on Applied Mathematics

The Calculus of Variations

Lecture 15

Multiple-Integral Problems Whose Lagrangian Functions Involve Derivatives of Higher Order than the First

We now suppose that L involves, in addition to the $N \times 1$ matrix $\begin{pmatrix} x \\ u \end{pmatrix}$ and the $1 \times n$ matrix u_x, the n-dimensional matrix u_{x*x}. This matrix is symmetric, and we could, if we so desired, confine our attention to those $\frac{n(n+1)}{2}$ of its elements which lie on and above its diagonal. The argument proceeds more easily, however, if we write L as a symmetric function of the symmetric matrix u_{x*x} as follows: Wherever we encounter any nondiagonal element $u_{x^p x^q}$, $p \neq q$, of u_{x*x} in L, we replace it by $\frac{1}{2}\left(u_{x^p x^q} + u_{x^q x^p}\right)$, and we define the n-dimensional matrix $L_{u_{x*x}}$ by regarding the n^2 elements of u_{x*x}, which we encounter in L when thus expressed as a symmetric function of u_{x*x}, as independent variables, ignoring the relations $u_{x^q x^p} = u_{x^p x^q}$, $q \neq p$,

which express the symmetry of the matrix u_{x*x}. Then the contribution to δL from the elements of u_{x*x} is the trace, i.e., the sum of the diagonal elements, of the product of the two n-dimensional matrices $L_{u_{x*x}}$ and δu_{x*x}, a typical element of this sum being

$$L_{u_{x^p x^q}} \delta u_{x^p x^q}, \quad p, q = 1, \ldots, n.$$

Assuming that the symmetric n-dimensional matrix $L_{u_{x*x}}$ possesses over $R + R$ continuous second derivatives with respect to the $n \times 1$ matrix x, we may integrate by parts each of the term $\int_{R + R'} \left(L_{u_{x^p x^q}} \delta u_{x^p x^q} \right) d(x)$ which appear in δI due to the dependence of L on u_{x*x}. Writing

$$\left(L_{u_{x^p x^q}} \delta u_{x^p} \right)_{x^q} = L_{u_{x^p x^q}} \delta u_{x^p x^q} + \left(L_{u_{x^p x^q}} \right)_{x^q} \delta u_{x^p}$$

$$\left\{ \left(L_{u_{x^p x^q}} \right)_{x^q} \delta u \right\}_{x^p} = \left(I_{u_{x^p x^q}} \right)_{x^q} \delta u_{x^p} + \left(L_{u_{x^p x^q}} \right)_{x^p x^q} \delta u$$

we have

$$L_{u_{x^p x^q}} \delta u_{x^p x^q} = \left(L_{u_{x^p x^q}} \delta u_{x^p} \right)_{x^q} - \left\{ \left(L_{u_{x^p x^q}} \right)_{x^q} \delta u \right\}_{x^p} + \left(L_{u_{x^p x^q}} \right)_{x^p x^q} \delta u$$

and the integrals of the first two terms on the right over $R + R'$ may be replaced by integrals over R'. Since δu is, by hypothesis, zero over R', the second of these two integrals over R' is zero. Similarly, the first of the two integrals over R' would be zero if we assumed that the $1 \times n$ matrix δu_x were zero over R', but since, after we have performed the integration by parts, we have to sum with respect to p and q, it is not necessary, in order to assure that this integral over R' be zero, to make such a strong assumption concerning the $1 \times n$ matrix δu_x. If $n = 2$, so that R' is a plane curve, the integral with which we are concerned is the line integral

$$\int_{R'} \left\{ \left(L_{u_{xx}} \delta u_x + L_{u_{yx}} \delta u_y \right) dy - \left(L_{u_{xy}} \delta u_x + L_{u_{yy}} \delta u_y \right) dx \right\};\ x = x^1,\ y = x^2$$

and it suffices to assume that the integrand of this line integral is zero, i.e., that the product of the 2×1 matrix

$$\begin{pmatrix} L_{u_{xx}} dy - L_{u_{xy}} dx \\ L_{u_{yx}} dy - L_{u_{yy}} dx \end{pmatrix}$$ by the 1×2 matrix, $(\delta u_x, \delta u_y)$, is zero

over R'. When $n = 3$, so that R' is a surface in 3-dimensional

space, the integral with which we are concerned is the surface integral

$$\int_{R'} \left\{ \left(L_{u_{xx}} \delta u_x + L_{u_{yx}} \delta u_y + L_{u_{zx}} \delta u_z \right) d\,(y,\ z) + \left(L_{u_{xy}} \delta u_x + L_{u_{yy}} \delta u_y + L_{u_{zy}} \delta u_z \right) d\,(z,\ x) + \left(L_{u_{xz}} \delta u_x + L_{u_{yz}} \delta u_y + L_{u_{zz}} \delta u_z \right) d(x,\ y) \right\} ;$$

$$x = x^1,\ y = x^2,\ z = x^3$$

and it suffices to assume that the product of the 3×1 matrix

$$\begin{pmatrix} L_{u_{xx}} d(y,\ z) + L_{u_{xy}} d(z,\ x) + L_{u_{xz}} d(x,\ y) \\ L_{u_{yx}} d(y,\ z) + L_{u_{yy}} d(z,\ x) + L_{u_{yz}} d(x,\ y) \\ L_{u_{zx}} d(y,\ z) + L_{u_{zy}} d(z,\ x) + L_{u_{zz}} d(x,\ y) \end{pmatrix}$$

by the 1×3 matrix $(\delta u_x,\ \delta u_y,\ \delta u_z)$ is zero over R'; and so on.

Making these assumptions, namely,

1) u is prescribed over R'.

2) The product of the $n \times 1$ matrix $L_{u_{x*x}}$ dS by the variation of the gradient $1 \times n$ matrix u_x of u is zero over R',

we have

$$\delta I = \int_{R+R'} \left[\left\{ \sum_{p,q \atop 1}^{n} \left(L_{u_{x^p x^q}} \right)_{x^p x^q} - \sum_{j=1}^{n} \left(L_{u_{x^j}} \right)_{x^j} + L_u \right\} \delta u \right] d(x)$$

and the Euler-Lagrange equation is

$$\sum_{p,q \atop 1}^{n} \left(L_{u_{x^p x^q}} \right)_{x^p x^q} - \sum_{j=1}^{n} \left(L_{u_{x^j}} \right)_{x^j} + L_u = 0 \ .$$

We have previously denoted the simple sum, $\sum_{j=1}^{n} \left(L_{u_{x^j}} \right)_{x^j}$,

by $\operatorname{div} L_{u_x}$, and we now denote the double sum, $\sum_{p,q \atop 1}^{n} \left(L_{u_{x^p x^q}} \right)_{x^p x^q}$,

by $(\operatorname{div})^2 L_{u_{x*x}}$. In this notation, the Euler-Lagrange equation of

of our problem of Type 2 of the calculus of variations appears as

$$(\operatorname{div})^2 L_{u_{x*x}} - \operatorname{div} L_{u_x} + L_u = 0 \ .$$

An important example of the theory is that for which

$L = (\Delta_2 u)^2 - 2 g u$, where $\Delta_2 u = \operatorname{Tr} u_{x*x}$ and g is continuous over $R + R'$.

Here $L_{u_{x*x}} = 2 \Delta_2 u E_n$, where E_n is the n-dimensional identity

matrix and L_{u_x} is the zero $n \times 1$ matrix. Since $(\mathrm{div})^2 \Delta_2 u$ is

$$\Delta_4 u = \sum_{p=1}^{n} u_{x_p x_p x_p x_p} + 2 \sum_{\substack{p<q \\ 1}}^{n} u_{x_p x_p x_q x_q}\,,$$ i.e., the symbolic square

of $\Delta_2 u$, the Euler-Lagrange equation is

$$\Delta_4 u = g$$

a partial differential equation of the fourth order; when $g \equiv 0$, this equation is known as the biharmonic equation. Since $L_{u_{x^* x}}$ is a multiple of the n-dimensional identity matrix, the second of our two assumptions listed above states that the product of the $n \times 1$ matrix dS by δu_x is zero over R' or, equivalently, that the normal directional derivative u_n of u is prescribed over R'; indeed dS is a multiple of u_{x^*} and $\delta u_x u_{x^*} = \frac{1}{2} \delta (u_x u_{x^*})$, so that $u_x u_{x^*}$ is prescribed over S and $u_x u_{x^*} = u_n^2$.

By the same argument as in the case where L was a function of X and u_x, not involving derivatives of u of higher order than the first, we see that if our problem of Type 2 of the calculus of variations is subject to a constraint of the type, $\int_{R+R'} \phi(u)\, d(x) =$ constant,

then the Euler-Lagrange equation is

$$(\mathrm{div})^2 L_{u_{x*x}} - \mathrm{div}\, L_{u_x} + L_u = \lambda\, \phi_u$$

where λ is an undetermined constant. In particular, when the constraint is $\int_{R+R'} u^2\, d(x) = \text{constant}$, this equation is

$$(\mathrm{div})^2 L_{u_{x*x}} - \mathrm{div}\, L_{u_x} + L_u = 2\lambda u$$

and this reduces, when $L = (\Delta_2 u)^2$, to

$$\Delta_4 u = \lambda u$$

The values of λ for which this partial differential equation possesses solutions, not identically zero, over $R + R'$ which vanish, together with their normal derivatives, over R', are the characteristic numbers of this boundary-value problem, the corresponding functions u being the characteristic functions of the boundary-value problem. It is easy to see that all these characteristic numbers are positive; indeed, on multiplying the equation, $\Delta_4 u = \lambda u$, by u and integrating the resulting equation over $R + R'$, we obtain

$$\lambda \int_{R+R'} u^2 \, d(x) = \int_{R+R'} u \, \Delta_4 u \, d(x)$$

$$= \sum_{p=1}^{n} \int_{R+R'} u \, u_{x^p x^p x^p x^p} \, dx + 2 \sum_{\substack{1 \\ p<q}}^{n} \int_{R+R'} u \, u_{x^p x^p x^q x^q} \, d(x)$$

$$= - \sum_{p=1}^{n} \int_{R+R'} u_{x^p} \, u_{x^p x^p x^p} \, d(x) - 2 \sum_{\substack{1 \\ p<q}}^{n} \int_{R+R'} u_{x^p} \, u_{x^p x^q x^q} \, d(x)$$

since $u = 0$ over R'. Since $\sum_{p=1}^{n} u_{x^p} \, dS^p = 0$ over R', we have

$$\sum_{p=1}^{n} \int_{R+R'} u_{x^p} \, u_{x^p x^p x^p} \, d(x) = - \sum_{p=1}^{n} \int_{R+R'} (u_{x^p x^p})^2 \, d(x) \quad \text{and}$$

$$\sum_{\substack{1 \\ p<q}}^{n} \int_{R+R'} u_{x^p} \, u_{x^p x^q x^q} \, d(x) = - \sum_{\substack{1 \\ p<q}}^{n} \int_{R+R'} u_{x^p x^p} \, u_{x^q x^q} \, d(x) \quad \text{and so}$$

$$\lambda \int_{R+R'} u^2 \, d(x) = \int_{R+R'} \left\{ \sum_{p=1}^{n} (u_{x^p x^p})^2 + 2 \sum_{\substack{1 \\ p<q}}^{n} u_{x^p x^p} \, u_{x^q x^q} \right\} d(x)$$

$$= \int_{R+R'} (\Delta_2 u)^2 \, d(x) \; .$$

In other words: each characteristic number λ is the quotient of the integral $\int_{R+R'} (\Delta_2 u)^2 \, d(x)$ by the integral $\int_{R+R'} u^2 \, d(x)$, where

u is a nontrivial solution, over $R + R$, of the partial differential equation $\Delta_4 u = \lambda u$, which vanishes, together with its normal derivative, over R'.

Let us now assume the existence of a function $u_1 = u_1(x)$, not identically zero, which possesses over $R+R'$, a continuous Laplacian $\Delta_2 u$, and which, finally, is such that in the class of all such functions u, the quotient $Q = \dfrac{\int_{R+R'} (\Delta_2 u)^2 \, d(x)}{\int_{R+R'} u^2 \, d(x)}$ has an absolute minimum, m_1 say, when $u = u_1$. Then, u being any member of this class of functions, $\int_{R+R'} \left\{ (\Delta_2 u)^2 - m_1 u^2 \right\} d(x) \geqslant 0$, the equality holding when $u = u_1$. Writing $u = u_1 + s f$, the integral last written becomes a differentiable function of s whose derivative, at $s = 0$, must be zero so that

$$\int_{R + R'} \left\{ \Delta_2 u_1 \, \Delta_2 f - m_1 u_1 f \right\} d(x) = 0 \quad .$$

Since the normal derivative of f vanishes over R,

$$\int_{R+R'} (\Delta_2 u_1 \Delta_2 f)\, d(x) = -\sum_{p=1}^{n} \int_{R+R'} \left\{ (\Delta_2 u_1)_{x^p} f_{x^p} \right\} d(x) \text{ , and}$$

this is, in turn, since $f = 0$ over R', $= \int_{R+R'} \left\{ (\Delta_4 u_1) f \right\} d(x)$.

Hence $\int_{R+R'} \left\{ (\Delta_4 u_1 - m_1 u_1) f \right\} d(x) = 0$, and this implies, since f is arbitrary except that it possesses a continuous Laplacian over $R + R'$ and vanishes, together with its normal derivative over R', that $\Delta_4 u_1 - m_1 u_1 = 0$ over $R + R$ so that m_1 is one of the characteristic numbers of our boundary value problem If λ is any characteristic number, the quotient Q for any characteristic function which corresponds to λ is λ, and so $\lambda \geqslant m_1$; thus m_1 is the least characteristic number, and we may use the Rayleigh-quotient method or the Rayleigh-Ritz modification of this method to approximate this least characteristic number.

In order to deal with characteristic numbers other than the least, let us assume the existence of a function u_2, not identically zero possessing a continuous Laplacian over $R + R$ and vanishing, together with its normal derivative, over R', and which is such that

$\int_{R+R'} u_1 u_2 d(x) = 0$ and, finally, which has the property that, in the class of all such functions,

$$Q = \frac{\int_{R+R'} (\Delta_2 u)^2 d(x)}{\int_{R+R'} u^2 d(x)}$$ has an absolute minimum, m_2, say,

when $u = u_2$. Then, u being any member of this class of functions,

$\int_{R+R'} \left\{ (\Delta_2 u)^2 - m_2 u^2 \right\} d(x) \geqslant 0$, the equality being valid when $u = u_2$. If f is any function possessing a continuous Laplacian and vanishing, together with its normal derivative, over R', we may determine the constant k such that $\int_{R+R'} u_1 g\, d(x) = 0$, where $g = f + k u_1$, k being, in fact, the negative of the quotient of the integral $\int_{R+R'} u_1 f\, d(x)$ by the integral $\int_{R+R'} u_1^2 d(x)$.

Writing $u = u_2 + sg$, so that $\int_{R+R'} u_1 u\, d(x) = 0$, no matter what the value of s, we find, as before, that

$$\int_{R+R} \left\{ \Delta_2 u_2 \Delta_2 g - m_2 u_2 g \right\} d(x) = 0 .$$

Writing $g = f + k\,u_1$, the coefficient of k in $\int_{R+R'} \left\langle \Delta_2 u_2 \Delta_2 g - m_2 u_2 g \right\rangle d(x)$

is easily seen to be zero; indeed, this coefficient is

$\int_{R+R'} (\Delta_2 u_2 \Delta_2 u_1 - m_2 u_2 u_1)\, d(x)$, which is the same as

$\int_{R+R'} (\Delta_2 u_2 \Delta_2 u_1)\, d(x)$, since $\int_{R+R'} u_2 u_1\, d(x) = 0$; finally,

$\int_{R+R'} (\Delta_2 u_2 \Delta_2 u_1)\, d(x) = \int_{R+R'} u_2 \Delta_4 u_1\, d(x)$, since u_2 vanishes,

together with its normal derivative over R', and this last integral is

zero, since $\Delta_4 u_1 = m_1 u_1$ over $R + R'$. Hence

$\int_{R+R'} \left\langle \Delta_2 u_2 \Delta_2 f - m_2 u_2 f \right\rangle d(x) = 0$, and the same argument as before

shows that this implies that $\Delta_4 u_2 = m_2 u_2$, so that m_2 is a characteristic

number. If λ is any characteristic number $> m_1$, with the associated

characteristic function u, the relation $\int_{R+R'} u_1 u\, d(x) = 0$ must hold;

indeed $\int_{R+R'} (\Delta_2 u_1 \Delta_2 u)\, d(x) = \int_{R+R'} (u_1 \Delta_4 u)\, d(x) = \lambda \int_{R+R'} u_1 u\, d(x)$

and similarly, $\int_{R+R'} (\Delta_2 u_1 \Delta_2 u)\, d(x) = \int_{R+R'} (u_2 \Delta_4 u_1)\, d(x) = m_1 \int_{R+R'} u_1 u\, d(x)$.

Since $\lambda > m_1$, the equality $\lambda \int_{R+R'} u_1 u \, d(x) = m_1 \int_{R+R'} u_1 u \, d(x)$

implies that $\int_{R+R'} u_1 u \, d(x) = 0$. Hence the quotient Q for $u \geqslant m_2$

so that $\lambda \geqslant m_2$, since $Q = \lambda$ when u is a characteristic function

associated with the characteristic number λ. Proceeding in this way,

we assume the existence of a function u_3, not identically zero, and

possessing the following properties:

1) $\Delta_2 u_3$ exists and is continuous over $R + R'$.

2) u_3 and $(u_3)_n$ vanish over R'.

3) Both $\int_{R+R'} u_1 u_3 \, d(x)$ and $\int_{R+R'} u_2 u_3 \, d(x) = 0$

4) $$Q = \frac{\int_{R+R'} (\Delta_2 u)^2 \, d(x)}{\int_{R+R'} u^2 \, d(x)}$$ has, in the class of all functions

possessing properties 1), 2), and 3), an absolute minimum m_3,

say, when $u = u_3$. Then m_3 is a characteristic number of our boundary-value problem, and if λ is any characteristic number $> \lambda_2$, $\lambda \geqslant m_3$.

Lecture 16

The Courant Maximum-Minimum Principle

We have seen that the least characteristic number, λ_1 of the boundary-value problem

$$\Delta_2 u + \lambda u = 0 , \quad x \quad R+R' ; \quad u = 0 , \quad x \subset R'$$

is a lower bound of the various Rayleigh quotients

$$R\,(v) = \int_{R+R'} \Delta_1 v \; d(x) \Big/ \int_{R+R'} v^2 \; d(x) \quad ; \quad \Delta_1 v = v_x \, v_{x*}$$

formed for all functions v which possess continuous gradients over $R+R'$ and which vanish over R'. If u_1 is a characteristic function corresponding to λ_1, $\int_{R+R'} \Delta_1 u_1 \; d(x) = - \int_{R+R'} u_1 \Delta_2 u_1 \; d(x)$, since u_1 is zero over R', and so $\int_{R+R'} \Delta_1 u_1 \; d(x) = \lambda_1 \; {}_{R+R'} u_1^2 \; d(x)$.

Hence, $R\,(u_1) = \lambda_1$, so that λ_1 is the greatest lower bound of the various Rayleigh quotients $R(v)$ formed for all functions v which

possess continuous gradients over $R+R'$ and which vanish over R'. By the same argument as that used at the end of the preceding lecture when discussing the boundary-value problem

$$\Delta_4 u + \lambda u = 0, \quad x \subset R+R' \; ; \; u = 0, \; u_n = 0, \quad x \subset R'$$

we see that the next larger characteristic number λ_2 is the greatest lower bound of the Rayleigh quotients $R(v)$ formed for all functions v which possess continuous gradients over $R+R'$ and vanish over R', and which, in addition are such that $\int_{R+R'} u_1 v \, d(x) = 0$. Similarly, the characteristic number λ_3 which is the next larger to λ_2 is the greatest lower bound of the Rayleigh quotients $R(v)$ formed for all functions v which possess continuous gradients over $R+R'$ and vanish over R', and which, in addition are such that both of the relations $\int_{R+R'} u_1 v \, d(x) = 0$ and $\int_{R+R'} u_2 v \, d(x) = 0$, where u_1 and u_2 are characteristic functions corresponding, respectively, to the characteristic numbers λ_1 and λ_2, hold; and so on.

The various characteristic functions $u_1, u_2, \ldots,$

corresponding, respectively, to the different characteristic numbers λ_1, λ_2, ... satisfy the relations:

$$\int_{R+R'} u_1 u_2 \, d(x) = 0 \; ; \quad \int_{R+R'} u_1 u_3 \, d(x) = 0 \; ; \; . \; . \; .$$

$$\int_{R+R'} u_2 u_3 \, d(x) = 0 \; ; \; . \; . \; .$$

and there is no lack of generality, since each of these characteristic functions may be multiplied by any constant, other than zero, in assuming that $\int_{R+R'} u_1^2 \, d(x) = 1$; $\int_{R+R'} u_2^2 \, d(x) = 1$, ..., and we shall do this. Let $v = c^1 u_1(x) + c^2 u_2(x)$ be any linear combination of u_1 and u_2 ; then $\int_{R+R'} v^2 \, d(x) = (c^1)^2 + (c^2)^2$.

To determine the Rayleigh quotient for v we observe that

$$\Delta_1 v = (c^1)^2 \Delta_1 u_1 + 2c^1 c^2 (u_1)_x (u_2)_{x*} + (c^2)^2 \Delta_1 u_2 \quad .$$

The integral of $(u_1)_x (u_2)_{x*}$ over $R+R'$ is the negative of the integral of $u_1 \Delta_2 u_2$ over $R+R'$, since u_1 vanishes over R', and this is the product by λ_2 of the integral of $u_1 u_2$ over $R+R'$.

We know that this latter integral is zero, and so

$$\int_{R+R'} \Delta_1 v \, d(x) = (c^1)^2 \int_{R+R'} \Delta_1 u_1 \, d(x) + (c^2)^2 \int_{R+R'} \Delta_1 u_2 \, d(x)$$

$$= (c^1)^2 \lambda_1 + (c^2)^2 \lambda_2, \text{ since}$$

$$\int_{R+R'} u_1^2 \, d(x) = 1, \quad \int_{R+R'} u_2^2 \, d(x) = 1$$

Hence $R(v) = \dfrac{(c^1)^2 \lambda_1 + (c^2)^2 \lambda_2}{(c^1)^2 + (c^2)^2}$, which is $\leqslant \lambda_2$, since $\lambda_1 < \lambda_2$.

Similarly, the Rayleigh quotient for any linear combination of u_1, u_2, u_3 is $\leqslant \lambda_3$, and so on. The useful feature of this result is that all the characteristic numbers $\lambda_2, \lambda_3, \ldots$, other than the least, λ_1, play the role of upper bounds rather than, as previously, lower bounds.

Let us now consider any function w which is integrable over $R+R'$, it being not required that w possess a continuous gradient over $R+R'$ nor that it vanish over R', and let us form the Rayleigh quotients for those functions v which possess continuous gradients over $R+R'$ and vanish over R', and, in addition,

are such that $\int_{R+R'} w\, v\, d(x) = 0$. Among such functions v is a linear combination, $c^1 u_1 + c^2 u_2$, of u_1 and u_2, the ratio of c^1 to c^2 being determined by the equation $c^1 \int_{R+R'} wu_1 d(x) + c^2 \int_{R+R'} wu_2 d(x) = 0$, unless both $\int_{R+R'} wu_1 d(x)$ and $\int_{R+R'} wu_2 d(x) = 0$, in which case c^1 and c^2 may be arbitrarily chosen. The minimum of these Rayleigh quotients, which minimum depends on w, is, accordingly, $\leqslant \lambda_2$ no matter what the function w. We have already seen that if $w = u_1$, this minimum is λ_2, the corresponding choice of v being u_2. Thus, λ_2 is the maximum, as w is varied, of the minimum of $R\,(v)$, formed for all functions v possessing continuous gradients over $R+R'$ and vanishing over R', and in addition, such that $\int_{R+R'} w\, v\, d(x) = 0$ where w is any function which is integrable over $R+R'$.

Similarly, if $w_1, w_2, \ldots, w_{p-1}$ is any set of p linearly independent functions which are integrable over $R+R'$, $p = 2, 3, \ldots$, the minimum of the Rayleigh quotients $R\,(v)$ formed for all functions v possessing continuous gradients over $R+R'$ and vanishing over R',

and in addition, such that the p-1 relations

$$\int_{R+R'} w_1 v \, d(x) = 0 \, ; \; . \; . \; . \qquad \int_{R+R'} w_{p-1} v \, d(x) = 0$$

hold, is $\leqslant \lambda_p$, and λ_p is the maximum of this minimum as the p-1 functions $w_1, \ldots, w_{p-1}$ are varied. This result is known as the maximum-minimum principle of Courant.

We now avail ourselves of the Courant maximum-minimum principle to show how to obtain, by means of the Rayleigh-Ritz method, approximations not only to the least characteristic number of the various boundary-value problems we have encountered but to the larger characteristic numbers as well. Let p be any positive integer, and let $v_1, \ldots, v_p$ be p linearly independent functions, each of which possesses a continuous gradient over R+R' and vanishes over R'. If v is the 1 x n matrix $(v_1, \ldots, v_p)$, and c is any constant p x 1 matrix, the matrix product vc possesses a continuous gradient over R+R' and vanishes over R'. The Rayleigh quotient for this matrix product is the quotient of two homogeneous quadratic forms in the p x 1 matrix c, and if we denote the matrices of the quadratic forms in the numerator and denominator of this quotient by M and N, respectively, we have

$$R\,(vc) = \frac{c^{*}\,M\,c}{c^{*}\,N\,c}$$

We seek now not merely the absolute minimum of $R\,(vc)$, but the various stationary values of $R\,(vc)$. Multiplication of the $p \times 1$ matrix c by any nonzero constant leaves $R\,(vc)$ unaffected, and we may determine, since $c^{*}\,N\,c > 0$, this multiplicative constant so that $c^{*}\,N\,c = 1$. Thus the stationary values of $R\,(vc)$ are those of the quadratic form $c^{*}\,M\,c$, the $p \times 1$ matrix c being supposed subjected to the constraint $c^{*}\,N\,c = 1$. In other words, at any stationary point of $R\,(vc)$, the linear form $d\,c^{*}\,M\,c$ in $d\,c$ must be zero for every $p \times 1$ matrix $d\,c$ which satisfies the relation $d\,c^{*}\,N\,c = 0$. This implies the existence of a number μ such that $M\,c = \mu\,N\,c$, μ being any one of the zeros of the polynomial in μ of degree p, $\det\,(M - \mu\,N)$. What we propose to show is that not only is the least of these zeros, μ_1 say, $\geq \lambda_1$, but the next greater of these zeros, μ_2 say, is $\geq \lambda_2$ and so on to the greatest of the zeros, μ_p, which is $\geq \lambda_p$. Our first step is to show that there is no lack of generality in taking N to be the p-dimensional identity matrix. Indeed, if we write $v' = v\,C$, where C is any non-singular p-dimensional matrix, the relation $v'c' = v\,c$ implies,

since the p elements of v are linearly independent, that $c = C\,c'$, so that

$$R(vc) = \frac{c'^{*} M' c'}{c'^{*} N' c'}\,, \text{ where } M' = C^{*} M\, C \text{ and } N' = C^{*} N\, C\,.$$

Then $\det(M' - \mu N') = (\det C)^2 \det(M - \mu N)$, so that $\det(M' - \mu N')$ is zero when, and only when, $\det(M - \mu N) = 0$. Since N is positively definite, we can determine a rotation matrix R such that $R^{*} N\, R$ is a diagonal matrix D with positive elements. Writing $C = R\, D^{-1/2}$, $N' = C^{*} N\, C$ is the p-dimensional identity matrix E_p. Thus, $\det(M' - \mu E_p) = 0$, so that $\mu_1, \ldots, \mu_p$ are the characteristic numbers of the p-dimensional matrix M', and our task is to show that $\mu_k \geq \lambda_k$, $k = 1, \ldots, p$. The fact that N' is the p-dimensional identity matrix E_p implies that

$$\int_{R+R'} v'_j\, v'_k\, d(x) = 0 \text{ if } j \neq k\,, \text{ while } \int_{R+R'} (v'_k)^2\, d(x) = 1\,,\ k = 1, \ldots, p,$$

and these relations remain valid if v' is subjected to the transformation $v' \to v'' = v\,R$, where R is any p-dimensional rotation matrix. Under this transformation $M' \to M'' = R^{*}\, M'\, R$ and R may be so chosen that M'' is the diagonal matrix whose diagonal elements are $\mu_1, \ldots, \mu_p$. From the Courant maximum-minimum principle we

know that, if $2 \leqslant r \leqslant p$, λ_r is the maximum, as the functions $w^1, \ldots, w^{r-1}$ are varied, of the minima of the various Rayleigh quotients $R(u)$ formed for those functions u possessing continuous gradients over $R+R'$ and vanishing over R' which are such that

$$\int_{R+R'} w^1 u \, d(x) = 0 ; \quad \ldots \quad \int_{R+R'} w^{r-1} u \, d(x) = 0$$

To this class of functions u belong those members of the family $v c = 0$, for which $\alpha^1 c = 0, \ldots, \alpha^{r-1} c = 0$, where

$$\alpha^1 = \int_{R+R'} w^1 v \, d(x), \ \ldots \ , \ \alpha^{r-1} = \int_{R+R'} w^{r-1} v \, d(x)$$

Hence the minimum of $R(u) \leqslant \mu_1 (c^1)^2 + \ldots + \mu_p (c^p)^2$, where the $p \times 1$ matrix c is subjected not only to the quadratic constraint $(c^1)^2 + \ldots + (c^p)^2$ but also to the $r-1$ linear constraints $\alpha^1 c = 0, \ldots, \alpha^{r-1} c = 0$.

Treating first the case $r = 2$, let us examine the minimum of $\mu_1 (c^1)^2 + \ldots + \mu_p (c^p)^2$, where $c^* c = 1$ and $\alpha^* c = 0$, α being a given $p \times 1$ matrix. At any stationary point of $\mu_1 (c^1)^2 + \ldots + \mu_p (c^p)^2$, subject to these constraints, $\mu_j c^j = \beta c^j + \delta \alpha^j$, where β and δ are

undetermined multipliers, and, since $c^* c = 1$, $c^* \alpha = 0$,

$\beta = \sum_{j=1}^{p} \mu_j (c^j)^2$ is the value of the quadratic form $\mu_1 (c^1)^2 + \dots + \mu_p (c^p)^2$ at the stationary point in question. Denoting by D the diagonal p-dimensional matrix whose diagonal elements are $\mu_1, \dots, \mu_p$, we see, on eliminating c and δ from the equations $(D - \beta) c - \delta \alpha = 0$, $\alpha^* c = 0$ that the minimum of $\mu_1 (c^1)^2 + \dots + \mu_p (c^p)^2$, subject to the constraints $c^* c = 1$, $\alpha^* c = 0$, is the least zero of the polynomial, of degree $p-1$, in β:

$$P(\beta) = \det \begin{vmatrix} D - \beta E_p & \alpha \\ \alpha^* & 0 \end{vmatrix}$$

The value of this polynomial when $\beta = \mu_1$ is $(\alpha^1)^2 (\mu_2 - \mu_1) \dots (\mu_p - \mu_1)$, its value when $\mu = \mu_2$ is $(\alpha^2)^2 (\mu_1 - \mu_2)(\mu_3 - \mu_2) \dots (\mu_p - \mu_2)$, and so on. Thus $P(\mu_1) \geqslant 0$, $P(\mu_2) \leqslant 0$, ... The numbers $\mu_1, \dots, \mu_p$ are the stationary values of $\mu_1 (c^1)^2 + \dots + \mu_p (c^p)^2$ when the $p \times 1$ matrix c is subjected to the single constraint $c^* c = 1$, and so we have the following result:

When the $p \times 1$ matrix c is subjected not only to the quadratic constraint $c^* c = 1$ but also to an arbitrary linear constraint $\alpha^* c = 0$, the stationary values $\mu'_1, \ldots, \mu'_{p-1}$ of the quadratic form $\mu_1 (c^1)^2 + \ldots + \mu_p (c^p)^2$ interlace the previous stationary values $\mu_1, \ldots, \mu_p$:

$$\mu_1 \leqslant \mu'_1 \leqslant \mu_2 < \mu'_2 < \quad \ldots \quad \leqslant \mu_{p-1} \leqslant \mu'_{p-1} \leqslant \mu_p$$

In particular, the minimum of $R(u)$, which is $\leqslant \mu'_1$, is $\leqslant \mu_2$. Since this holds, no matter what the function w^1, it follows that λ_2, which is the maximum of this minimum as w^1 is varied, is $\leqslant \mu_2$.

When $r = 3$ we have two linear constraints $\alpha_1^* c = 0$, $\alpha_2^* c = 0$ instead of merely one. We use the first of these two linear constraints to write our Rayleigh quotient $R(vc)$ as a quadratic form in a $(p-1) \times 1$, rather than a $p \times 1$, matrix c'. The quadratic constraint $c^* c = 1$ appears as $c'^* A c' = 1$, and this may be written as $c''^* c'' = 1$ by setting $c' = B c''$, where B is an appropriately chosen nonsingular $(p-1)$ - dimensional matrix. Hence,

by the result for $r = 2$, the new stationary values, $\mu''_1, \dots, \mu''_{p-2}$, p-2 in number, interlace the immediately preceding stationary values, $\mu'_1, \dots, \mu'_{p-1}$:

$$\mu'_1 \leqslant \mu''_1 \leqslant \mu'_2 \leqslant \mu''_2 \leqslant \dots \leqslant \mu'_{p-2} \leqslant \mu''_{p-2} \leqslant \mu'_{p-1}.$$

Hence $\lambda_3 \leqslant \mu_3$, since $\mu'_2 \leqslant \mu_3$. Continuing in this way we obtain $\lambda_j \leqslant \mu_j$, $j = 1, \dots, p$.

Example We saw in Lecture 10, when discussing the boundary-value problem: $x_{tt} + \lambda x = 0$, $-1 \leqslant t \leqslant 1$, $x(-1) = 0 = x(1)$, that the minimum of $R(s) = \frac{3}{2}\left(\frac{35 + 14s + 11s^2}{21 + 6s + s^2}\right)$ is a good approximation to, being only slightly greater than, the least characteristic number, $\frac{\pi^2}{4}$, of this problem. The two stationary values of $R(s)$, one of which is the absolute minimum and the other of which is the absolute maximum of $R(s)$, are the zeros of the quadratic polynomial in μ

$$\det\begin{pmatrix} 105 - 42\mu & 21 - 6\mu \\ 21 - 6\mu & 33 - 2\mu \end{pmatrix} = 48(\mu^2 - 28\mu + 63)$$

The smaller of these two zeros, 2.4674 to 4 decimal places, is a good approximation to $\frac{\pi^2}{4}$, which is also 2.4674 to 4 decimal places. The larger of the two zeros, 25.5326 to 4 decimal places, is an approximation, in excess, to the second characteristic number, $9\frac{\pi^2}{4} = 22.2066$, but the approximation is not nearly so good as that furnished by the smaller of the two zeros.